Michael Schulte-Mecklenbeck

Tracing the Decision Maker

June, 2005

University of Fribourg
http://e-thesis.unifr.ch

Dissertation zur Erlangung der Doktorwürde an der Philosophischen Fakultät der Universität Freiburg (Schweiz).

Genehmigt von der Philosophischen Fakultät auf Antrag der Professoren
Prof. Dr. Oswald Huber (1. Gutachter) und
Prof. Dr. Anton Kühberger (2. Gutachter).

Freiburg, 18. Februar 2005. Prof. Dr. Richard Friedli, Dekan

Acknowledgements

I want to thank the following people who all contributed to this thesis either professionally or privately:

Oswald Huber for letting me pursuit my ideas and for the daily coffee at 10.00.

Anton Kühberger who has been an advisor, teacher and friend for many years now.

Michael for proof reading this thesis. You are now officially a member of the “Club of four” (requirement for membership is the reading of the whole thesis)!

Moritz for agreeing to write his master thesis about WebDiP and for standing my constant nagging about details.

Simon for opening the doors to Mathematics for me (at least a bit) – our discussions on Markov Models and sequential data analysis were really one of the fun parts of this thesis.

Katrin and Paul for a still growing friendship.

In no particular order my colleagues at the Department of Psychology, University of Fribourg (who became friends over the last four years) – Odilo, Elisabeth, Jörg, Ursina (c/o UC San Diego), Andreas, Andrea, Yves, Pascal and Eric.

Karoline – for so many things – first of all for accepting me as I am. For being there for me when there is need and especially for letting me go when I need some room. For helping me with the task generation which would have developed into a kind of nightmare otherwise (ok, it actually has developed into one). For coming to Switzerland and living with me. It is simply great to be with you!

Good introductions should be short – I think good acknowledgments follow the same rule.

Abstract

This thesis focuses on two themes: a methodological and a decision making one. The methodological part introduces the development of a computerized information search tool. The program, called WebDiP (Web DecisIon Processes), enables the researcher to track participants while they are searching for information in a database. An experiment can be done either in the laboratory or on the Web. The first study identifies methodological shortcomings in an earlier software version and clarifies these problems with the introduction of WebDiP.

The decision making part concentrates on the question whether the content (domain) of a decision has an influence on decision making, i.e., in the introduced process tracing approach, the information search. Three domains (business, law and medicine) are selected and for each domain two tasks developed. All six tasks (3 domains X 2 tasks per domain) have the same structure, probabilities and outcomes. In the second study a large scale online experiment with over 360 participants pursues the research question whether different domains elicit different search behaviors or differences in used information (type) are pursued. Results of this experiment illustrate problems within the assumption that one task is always representative for one domain. Furthermore, patterns are identified which are representative for a task but not necessarily for a domain.

Both themes are discussed in light of the current literature on online research, process tracing and domain differences.

Contents

List of Figures

List of Tables

1 Introduction

Let me start with three observations.

Psychological decision making research has its *roots in economics*. This is a fact that can still be realized when research tasks are inspected. Gambles, the favorite tool of an economist, are popular in Psychology, too. Most often these gambles are presented as a text which requires a reaction of the participant. This text includes all the information necessary for the participant to make a decision. The artificial nature of this approach is obvious – a situation in real life, where all information is given – hard to imagine.

The second observations concerns the *universal claim* that models in decision making are valid in all areas of life. But can this really be the case? Is it possible that human beings use one mode for decision making in every available life domain? Can it be that the content of a decision does not matter at all and that a description of the decision maker is possible in a universal sense? The answer, while seeming easy, yet opens up a multitude of questions. In a lot of other areas of Psychology, thinking about *content or domain-dependent behavior* is common – a central claim of this thesis will be that this should also be true for decision making research.

On a more general level the third observation is about *information need* or the *information society*, both expressions which have arisen with the rapid growth of the Internet over the past 20 years. It is on the one hand the amount of information available to the public and on the other hand the ease of access to this information through the usage of *the net* – virtually at everybody's fingertips. The easy way of access to an enormous amount of information makes the Internet a unique development of mankind. The common tool to find information is a search engine (large databases filled with searchable indexes of as many Web-sites as possible). Today the entering of a keyword into a search engine often results in more information (not necessarily the best one) than processable by a single user.

What do these three paragraphs have in common? First of all they are about a typical type of information presentation in risky decision making research – a text with all the necessary information for solving a task. They are about the question what influence the environment has on the representativeness of a decision – the domain. They are about a method by which we gain access to information today – search engines on the Internet. Concurrently they represent the two central themes of this thesis.

(1) The methodological theme: the development of an online information search tool that does not present all the necessary information as in a traditional decision making task but demands search action for gathering information by the participant. After the development of a computerized search version, this *offline* version was transferred to the Internet and validated. The first empirical study (Chapter 6, p. 53) addresses the differences between running an information search experiment in the laboratory and on the Web (experiment 1+2). The identified methodological

problems (mainly useability and database related problems) are discussed. A rebuild of the software named WebDiP – Web DecisIon Processes (Chapter 7, p. 67) and a test of the critical instances from the former research tool is provided (experiment 3). The result of this process is an information acquisition tool which enables the researcher to analyze how much and which information is gathered in what order. Main advantages of this approach are the more or less uninfluenced, automatic gathering of processing data while a decision is being made. A detailed methodological description can be found in the chapter mentioned above.

(2) The decision making, theoretical theme highlights the fact that decision making is indeed highly content and domain specific. Following a *domain differences* hypothesis, which was introduced in problem solving literature, an attempt at improving the understanding of information-search and information-demand in different domains is presented. Relevant literature is identified and clustered either into a *decision mode dependent* or a *decision domain dependent* group. The later group is then used in an information search study (experiment 4).

The following steps are taken to accomplish these two themes:

1. Literature overview: A brief overview on concepts in risky decision making research is given first. Then relevant literature for the subsequent experiments is located and discussed. The area of *process tracing* – from a methodological point of view – as well as theories using process tracing as a method are presented. Finally an overview on classification techniques and literature dealing with decision modes/domains is given.
2. Development of the necessary experimental tools: An approach first described in Huber, Wider, and Huber (1997) and labeled Active Information Search (AIS) will be used as the methodological basis. AIS in the traditional version lets the participants ask for information in order to solve a decision making problem (see Chapter 3.4, p. 16 for a detailed description of the method).
   The basic methodological ideas of AIS are used and transferred into an experiment run on a computer. The questions and answers are stored in a database which the participants access through different search methods. *Study one* (see p. 53) examines two of these search methods (list-search and keyword-search) and furthermore contrasts running the same study on the Web and in a laboratory. Differences between laboratory and Web as well as methodological problems of the computer approach are discussed (experiment 1 and 2).
3. Further development of the experiment's software: An enhanced version of the software used for the first study called WebDiP is introduced in Chapter 7, p. 67. Shortcomings in the search algorithm and programming of the user interface are overcome in the new version. An experiment demonstrating an increase in demanded information (a problem identified in experiment 1 and 2) and a more useable interface for WebDiP is presented (experiment 3).
4. Moving on from this theoretical basis, literature dealing with *decision modes* on the one hand and *decision domains* on the other is addressed (the distinction between the two is discussed in this chapter, too). In a *clusteranalysis*, the various decision modes are grouped and a taxonomy of decisions (concerning modes) is built. On the decision domain side (due to a lack of literature dealing with this topic) a survey in the JDM (Judgment and Decision Making) newsletter tried to identify decision domains relevant for experts, i.e., decision researchers.
5. Task generation: Tasks from different papers by Huber et al. (1997); Huber and Macho (2001); Rettinger and Hastie (2001) are considered and adopted. Additionally, the development of new tasks was necessary in order to have at least two information search tasks for every identified domain. Two tasks were generated because it is common practise to use one task as a representative for a domain, an issue that should be overcome by this step.
6. Implementation of an online experiment: Finally, the generated tasks (two from each of the following domains: business, law and medicine) were administered in an online experiment (see

Study 2, p. 77). Identification of information search types and detailed analysis of information acquisition behavior finalizes this last step (experiment 4).

Research questions

According to the outlined steps above the following research questions shall be answered:

- *Tool development.* Does a transfer from of an AIS experiment to the computer change the information search behavior? The AIS method includes the participants' questions which are answered by the experimenter through information on cards. These cards are given to the participant and stay with him until the end of the experiment (see a detailed description of the experimental method in Chapter 3.4, p. 16). In order to get a more objective experimental setting and an automated procedure for data collection the information on the cards from AIS are transferred into a computer database. Participants' are able to directly "question" the database within this computerized version.
- *Comparison between Web and laboratory.* The above described computerized version lends itself for usage over the Internet. Therefore the following question should be answered: Is an experiment conducted in the laboratory comparable to the same experiment run on the Web? There already exists a vast amount of literature comparing different methods of psychological research in the laboratory and on the Web (Batinic, Reips, & Bosnjak, 2002; Birnbaum, 2000; Reips, 2002). Especially questionnaires are often tested on validity and reliability issues (for a recent overview see Gosling, Vazire, Srivastava, & John, 2004). For psychological experiments in general, only a few studies are available for decision making experiments, this number is even smaller (e.g., Anderhub, Müller, & Schmidt, 2001; Birnbaum, 1999, 2004; Shavit, Sonsino, & Benzion, 2001). Because of this situation, it seems intriguing to combine the method of process tracing from AIS with a decision making task and compare these in two locations – the laboratory and the Web.
- *Decision making domains and modes.* Different content should yield different decision behavior. The influence of content on a decision is of central interest next. What categorizations of decisions are available? Due to the lack of generally accepted definitions of decision *domains* and decision *modes*, both are described, differentiated and discussed using examples from the literature.
- *Information need.* Are there differences in information need between domains? What effect does content have on the information search *process*? The introduced differentiation is used to get a clearer picture on the content's influence on the type and sequence of information a person is interested in. For this research question different tasks will be compared concerning their information need in a process tracing study.

Structure

The thesis is structured as follows: *Chapter 2* provides a brief overview on important concepts in the area of decision making research. *Chapter 3* reviews different process tracing methods with a focus on the Adaptive Decision Maker (ADM) paradigm. *Chapter 4* highlights theoretical approaches to decision making research which all have the equality of process tracing as a method. The second central theme of this thesis is introduced in *Chapter 5* when literature on decision domains/modes and a classification of these is presented.

The experimental part consists of three chapters. *Chapter 6* wants to demonstrate the usage of a database as a research tool in Web and laboratory based experiments. *Chapter 7* describes the development and technical details of WebDiP – a tool for running information search experiments on the Internet – and an experiment for comparing WebDiP to the former software version. Finally, an experiment using tasks from different decision domains (in an internet based WebDiP study) concludes *Chapter 8*.

2

Basics

This section wants to give a brief introduction into concepts used in decision making research. An overview on a topic like decision making is of course an endeavor worth a thesis on its own, nevertheless shall the mentioned concepts set a framework. After basic components of a decision situation are described, multi attribute decision making will be introduced. Some thoughts about risk end this chapter. The central theoretical ideas of this thesis are additionally outlined in more depth in Chapter 4, p. 21 – *Theoretical approaches* and Chapter 5, p. 41 – *Decision modes and decision domains*!

What is decision making?

When making a decision one tries to identify and choose between *alternatives* based on *subjective values* and *preferences*. A fundamental precondition for a decision is the *existence of alternatives* – without these no energy has to be invested into information search, building of preferences and retrieval of values. Another factor that is crucial within the decision making process and studied frequently over the last 50 years is *uncertainty*. Reducing uncertainty about alternatives is an important step within the decision making process. Uncertainty is reduced through *information search* about the alternatives, given that complete information gathering is hardly ever possible a certain risk is bound to a large number of decisions.

What are examples for a risky decision making task? As described above at least one of the alternatives has undesirable consequences, and the occurrence of these consequences is uncertain. For example, a patient is asked to decide whether he wants to take a drug that may have dangerous side effects or to undergo surgery. Undesirable consequences of the latter could be complications during surgery and the possibility that the surgery may not be successful. Other well-researched types of risky decisions are gambles, such as roulette, lotteries, and bets. Later it will be argued that these types of decisions are special in terms of information presentation however they are most frequently used in decision making research.

2.1 Components of a decision

- Information. General knowledge about a decision is as well incorporated into this component as the probability of an alternative, the probability about negative events or other specific information. An important point is that while substantial information is desirable, the statement that "the more information the better" is not necessarily true. Too much information can actually reduce the quality of a decision. See, e.g., the *fast and frugal* approach Gigerenzer, Todd, and ABC Research Group (1999) for prominent usage of the "less is more" idea.

- Alternatives. These are the possibilities one has to choose from. Possibilities exist in various form, they can either be objects or actions which are most often goal orientated. But also more complex possibilities exist like strategies (for longer periods) or rules (how to do something). Alternatives can be identified (that is, searched for and located) or developed (created during the decision making/information search process).
- Attributes. These are the characteristics or requirements that each alternative must possess to a lesser or greater extent. Usually the alternatives are rated on how well they possess each attribute. For example, alternative Mercedes ranks an 8 (out of 10) on the attribute of luxury, while alternative 2CV only ranks 2 on the same attribute.
- Goals. Goals are long term objectives somebody wants to reach. Often decision makers collect a number of alternatives (cars to buy or universities to go to) and then ask "Which should I choose?". This is done without thinking of goals at first hand. Goals are more or less abstract "being successful" or "saving money" are two examples. Goals can be divided into subgoals in the case of "being successful" subgoals are, e.g., earning a sufficient amount of money or having an interesting job.
- Preferences. These reflect an internal hierarchy of the decision maker. Some people prefer excitement to calmness or certainty to risk or quality to quantity. Thus, when one person chooses to ride the roller coaster in an amusement park and another chooses a ride on a more laid back attraction, both may be making good decisions, based on their individual preferences. Preferences can be measured through asking for the favorite out of two or more alternative or through a person's ranking of such alternatives.

2.2 Multi attribute decision making

Central to most decision situations is the existence of many attributes that have to be evaluated according to the decision makers preferences. Inherent with such a multi attributes situation is the *tradeoff* people have to make, e.g., between costs and quality or effort and accuracy (a distinction which is central to the adaptive decision maker approach introduced in Chapter 4.3, p. 28).

Six assumptions about characteristics of multi attribute decision strategies are illustrated now.

1. A *compensatory strategy* allows a good value on one attribute to make up for bad values on other attributes. Whereas a *noncompensatory strategy* makes no tradeoffs – a bad value on an important attribute ensures the skipping of this attribute.
2. The question whether the same amount of information is processed for both alternatives is addressed in the distinction between *consistent* and *selective processing*. Consistent processing involves examination of all information from all alternatives. However the more common strategy seems to be the selecting of interesting parts of information for further considerations in the decision process.
3. More generally speaking, the *amount of processing* is either reduced to a selection of information with the risk of ignoring important information or the attempt of processing all of the given information.
4. *Alternative-based versus attribute-based processing* plays a central role in the adaptive decision maker approach. The two dimensions describe the *path* a decision maker paces when searching for information. Payne (1976) proposed an index (Payne Index, PI, see p. 37 for details) to describe this behavior.
5. Strategies differ if (or if not) an evaluation of the alternatives is performed. This *formation of evaluation* of groups is a central distinction for many decision rules.
6. Finally, the *quantitative versus qualitative reasoning* characteristic groups strategies into, e.g., ones that count or calculate certain aspects versus strategies that rate values or utilities.

2.3 The perception of risk

The discussion of risk in the decision making literature is very broad and goes back to the roots and the very beginning of this research field. Especially normative models like, e.g., expected utility theory include risk components through usage of probabilities of desirable events. However there is also a large literature referring to risk in a more applied setting like the *perception* and the *communication of risk* (see Slovic, 2001 for an excellent introduction).

Risk is more than a statistic telling us that a male smoker who inhales reduced his life expectancy by 8.6 years whereas a female smoker who inhales looses 4.6 years (Cohen & Lee, 1979, as cite in Plous, 1993). Following this statistic – should a female smoker be half as worried as a male smoker is? Do people judge the risk of smoking – a "voluntary" risk – equal to the risk of, e.g., an electrical power plant – an "involuntary" risk? Both questions can safely be answered with *no*.

Slovic (1987) describes three dimensions corresponding to risk perception in the public. The first dimension is *dread risk* which is characterized by a perceived loss of control, fatal consequences and inequitable distribution of risks and benefits. The most extreme examples for such a risk is nuclear power (in the days these dimensions were described) or nowadays most often terror attacks. The second dimension was named *unknown risks* which are unobservable and often delayed in their "manifestation of harm". The usage of genetics for the improvement of food is an example for this class. The final third dimension includes the number of people threatened by a risk, i.e., an exposure a base rate. The author demonstrated in various studies that these three factors could be replicated over and over again.

2.4 Summary

The above described components of a decision combined in a multi-attribute decision task with risky alternatives are also used in this thesis. However the above examples for risk perception with very important events (including life and death) can not be used in such a realistic manner here. Still the notion that *hypothetical* tasks bear a certain connection to *real* tasks holds but there are differences when a participant searches for information concerning a task which is purely hypothetical and which has no immediate influence on the participant's health or life. Despite this alleged drawback studying the later introduced approach enables the researcher to access stages in the decision making process which would be rather inaccessible given, e.g., the life of a participant is currently threatened.

3

Following a trace

H. Montgomery and Svenson write in their introduction on “Human Decision Making and Process Tracing” that there...

> “...has been a growing interest in the cognitive and evaluative processes behind a particular choice or judgement.”
> (H. Montgomery & Svenson, 1989a, p. XI)

Although this quote is now more than 15 years old, it still accurately describes the aim of process tracing research. The differentiation between what is happening while the decision maker is *in* a task and what is happening *after* a task is also addressed in the following classification of models in the decision making literature. Two groups can be identified: *structural-* and *process-models*. (1) *Structural models* – have an input[1] and an output[2]. The main issue of such models is the analysis of the second part – the *postdecisional* outcomes. Therefore it is difficult to gain insight into the processes going on during a task. (2) *Process models* overcome this drawback by observing traces a decision maker leaves, actively or passively, while accomplishing a task. The order and the amount of acquired information as well as its representation is of central interest.

In a meta-analysis Ford, Schmitt, Schechtmann, Hults, and Doherty (1989) gave an overview of two process tracing methods (verbal protocols and information boards). First, the authors made an important distinction between *compensatory* and *non-compensatory* strategies (see Chapter 2.2, p. 6) which describe the order in which information is searched. In a *compensatory* strategy a compensation of a low value in one attribute (cue, dimension[3]) is possible through a high value in another. Contrary to this, with a *non-compensatory* strategy a low value in one attribute can not be compensated. When choosing between different apartments (alternatives, options), low price could compensate for few rooms. However if the next school should be within 10 minutes no compensation through an additional garden would be possible (see Chapter 4.1, p. 21 for empirical results of strategy change during the information search process and Chapter 4.3, p. 28 for further discussion).

In the reviewed literature Ford et al. (1989) identified non-compensatory strategies more often than compensatory ones. Task complexity played a moderating role to this result – the smaller the number of attributes and alternatives – the more a compensatory strategy was used and vice versa. Ford et al. (1989) predicted a *two step process* with eliminating alternatives and attributes (non-compensatory) before switching to a compensatory strategy for the rest of the task.

---

[1] which is in most cases a task text, however other inputs could be, e.g., attributes and alternatives in a MAUT like decision – if there is no process tracing involved
[2] the reaction of the decision maker
[3] further on *alternative* and *attribute* will be used following Payne, Bettman, and Johnson (1993)

In addition to this two step model several critical remarks on process tracing research are made by Ford et al. (1989). Here, only two that are relevant for this work will be mentioned. (1) The *effects of a person and the environment* are often ignored in favor of comparing different tasks within a process tracing framework. The authors suggest that one person who is put into a different task situation would act differently and conclude that ...

> ... the large amount of research using process tracing to study task effects had not led (with the possible exception of task complexity) to the cumulation of knowledge necessary to begin the development of a taxonomy of task characteristics which have been found to have predictable effects on strategy use. (Ford et al., 1989, p. 104)

This issue is addressed in experiment 4 (see Chapter 8, p. 77). (2) A second point of criticism concerns the *labeling of alternatives* in a decision matrix (decision/ information board). The predefined names (in a decision board alternatives and attributes are presented readily named to the decision maker) must be accepted by the decision maker – no studies dealing with unlabeled alternatives were found in the review. Nevertheless the effect of presenting labeled alternatives to the decision maker is unclear. WebDiP, which will be introduced later (Chapter 7, p. 67), solves the above issues.

Another meta analysis dealing with the aim of overcoming the simple structural model by building process models was introduced by Harte, Westenberg, and Someren (1994). The authors analyzed journal papers between 1976 and 1993 about individual decision makers in a multi-attribute decision problem with thinking aloud protocols. Twenty three papers were classified into three groups: *non-sequence*, *sequence* and *process models*. The *non-sequence* category equals the above described structural model, whereas papers in the *sequence* category at least make predictions about the order in which processes take place. The *process model* category gives a description of the full process in a decision making task. Only four studies were included into the process model class – Payne, Braunstein, and Carroll, 1978; Einhorn, Kleinmuntz, and Kleinmuntz, 1979; Svenson, 1985; Johnson, Jamal, and Berryman, 1991, as cited in Harte et al. (1994). It is astonishing that although there was a large number of studies segmenting the thinking aloud protocols into single statements – a process model generation was only suggested in four.

What are the components of such a process model? For the construction seven steps are distinguished by Harte et al. (1994):

1. Task analysis – how can the task be performed
2. Construction of a psychological model – a prediction of how people will behave
3. Construction of a coding schema – describes how the model's predicted processes do appear in the text
4. Collection of protocols
5. Segmentation of protocols – dividing protocols into small parts
6. Coding of protocols – applying the coding schema
7. Comparison of protocols with predictions

Because of the computerized research approach, introduced later, certain parts of this description are automated and therefore of only secondary interest in this thesis. Nevertheless steps 1–3 as well as 6 and 7 seem applicable. The following paragraphs describe the suggested action for each step by Harte et al. (1994).

Step 1 *Task analysis*: Four strategies a participant can choose from when performing a multi-attribute decision problem are distinguished: (1) *linear integration* of the aspects of each alternative into an overall value, (2) *conjunctive* applying a minimum threshold and eliminating alternatives beneath it, (3) *elimination by aspect* – starting with the most important attribute and eliminating all alternatives that do not meet a minimum threshold and (4) *additive difference* – evaluation of differences between alternatives and subsequently summing up.

Step 2 *Construction of a psychological model*: This step contains the application of general psychological theories to given decision models. Harte et al. (1994) concentrate on results of working memory research about information storage. With these limitations in mind (e.g., only limited amount of information can be stored in working memory see e.g., Baddeley, 2003 for a recent overview), the application of a linear decision strategy seems questionable, because it has to result in an information overload. Too much information has to be stored before a decision can be made.

Therefore, Harte et al. conclude that participants should make use of heuristics. All of the discussed heuristics (e.g., elimination by aspect ...) make concessions to information retrieval but nevertheless perform remarkably good when it comes to simulations or experimental studies (for an elaborate discussion on the performance of heuristics see Chapter 4.3, p. 28).

Step 3 *Construction of a coding schema*: In thinking aloud studies, it is important to assure the connection between the theoretical assumptions and the fragments of the text in the protocols[4].

Step 4 – 6 *Collection, segmentation and coding of protocols*: In these steps the actual data is collected. The process data are recorded, segmented into small parts and coded according to a coding schema. The development of this coding schema is most often done parallel and iteratively with the segmentation process.

Step 7 *Comparison of protocols with predictions*: A match of the protocols with the model would result in acceptance of the model and the connected predictions. However, Harte et al. (1994) state that this match is often hard to achieve and even harder to quantify.

Different methods were generated for data collection in process tracing research. They can be divided into the following groups and are introduced in the subsequent sections. The first group deals with *introspection* and *thinking aloud protocols*. The second group includes *information boards* (here eye-tracking will be introduced, too) and other *computerized methods*. The third group includes *online research* and process tracing. Within this group, the method of *Active Information Search* (AIS) although being a laboratory method is introduced, too, because it serves as the basic idea of the online experiments in Study 1 (see Chapter 6, p. 53) and Study 2 (see Chapter 8, p. 77).

3.1 Thinking aloud and introspection

Introspection is one of the first methods used in experimental Psychology (e.g., Ebbinghaus, 1885, as cited in Baddeley, 1990). Participants are trained to focus on a specific process and report on it. Additionally, theorizing of participants on the observed process is required – everything the participant could comment on the observation was of interest. Today introspection is completely replaced by other methods (e.g., thinking aloud, see below). Today's research standards can not be guaranteed with this approach!

In *thinking aloud* or *verbal protocol* studies, participants who are naive to the subject of research are asked to verbalize anything that comes to mind (Ericsson & Simon, 1993). No training is required before such an experiment. The verbalizations of participants are recorded and transcribed. During the transcription process, statements are segmented into "meaningful components" (these are defined as everything between pauses generated during the speaking process). After the segmentation, data is coded according to a predefined coding schema. At least two independent

[4] This coding is also central when using a computerized approach, because results can be influenced considerably by changes in the coding of information.

coders should perform this task. An analysis of intercoder reliability can post-determine ambiguous codings and result in a reconstruction of the coding schema.

Verbal protocol studies have a tradition going back more than 80 years now. Watson (1920, as cited in Svenson, 1989) already preferred the thinking aloud method to introspection which he classified as "unscientific". Two recording times of thinking aloud protocols can be differentiated (1) *concurrent thinking aloud protocols* ask the participant to speak while doing a task. The dual task characteristic of concentrating on a task while speaking about it was often criticized for demanding access to processes not accessible at that moment. (2) *Retrospective protocols* analyze the process after completion of the task. Problems that could arise with this method are connected to shortcomings in the participant's memory. A possible tendency of constructing a trustworthy story *after* the task is completed could decrease validity of the method, too (Nisbett & Wilson, 1977). A multitude of research questions was covered with thinking aloud protocols – three examples are given now – a framing study (Maule, 1989), a hypothetical business study (H. Montgomery & Svenson, 1989b) and a medical decision making study (Backlund, Skaner, Montgomery, Bring, & Strender, 2003).

Maule (1989) let participants think aloud in an Asian Disease like task (Tversky & Kahneman, 1981). The same choice patterns (risk-averse choices in the gain- versus risk-seeking choices in the loss-frame) like in the original experiment were found. A comparison of the verbal protocols between decision frames (gain and loss) in terms of, e.g., usage of gain and loss words between frames revealed the predicted relation of the gain-frame with a majority of positive words and the loss-frame with a majority of negative words. Unfortunately, this pattern was only found with approximately 50% of the participants. A mixed or opposite pattern was present in the other half. This experiment nicely demonstrates the move from the structural model to the process model. The outcomes/choices are consistent with predictions of the structural model. Taking the next step and inspecting the processes underneath this will reveal a much sketchier picture.

In a more naturalistic experiment, H. Montgomery and Svenson (1989b) let subjects choose between different houses while reporting the decision process. Of central interest for the authors was the attention subjects set to the alternatives during the decision process. At the beginning of the protocols (the decision processes), a clear majority of participants turned to the alternative (finally) not chosen. It is assumed that those participants first generated a hypothesis about an alternative, realized that they could not confirm their hypothesis and try to restore it by finding information confirming their initial idea (see Chapter 4.1, p. 21 for details on the theory).

A recent study by Backlund et al. (2003) focused on the possibility of replacing verbal-protocols with a rating scale approach in a medical decision making task. This comparison is interesting because of the greater effort (in terms of time and analysis) a thinking aloud study takes in comparison with a study using rating scales. On a theoretical level, the authors tested for applicability of two process tracing theories *Search for Dominance Structure* (SDS) and *Differentiation and Consolidation Theory* (DiffCon, see Chapter 4, p. 21 and p. 24 respectively). In the study participants (medical doctors) had to decide on a hypothetical case offering two alternative – prescribing a drug or not prescribing it. The information was presented on a computer on consecutive screens (compare this method to WebDiP, Chapter 7, p. 67, which presents information in "pieces", too). In terms of interest in the alternatives during the information search process, results showed that the introduction of rating scales as an additional measure did not change the thinking aloud process remarkably. The differentiation (as predicted by DiffCon Theory) between alternatives was found with both methods. However, results were much stronger for the thinking aloud group alone in comparison with the combined group (thinking aloud and rating scale). Backlund et al. (2003) interpreted this result as a valid hint for the possible replacement of the time-consuming thinking aloud protocols with relatively simple rating scales.

In a more general comment, Harte and Koele (2001) criticize the verbal protocol method for being more time-consuming, too. They highlight discussions on *reactivity* (i.e., the process of verbalization during the task changes the task itself) and *nonveridicality* (i.e., a not representative reflection a model gives of an underlying process) of the method (Russo, Johnson & Stephens, 1989, as cited in Harte & Koele, 2001). The reactivity problem was addressed in Ericsson and Simon (1993) with the conclusion of a slower process if thinking aloud is performed but no change in the basic sequence of thoughts during the process. The same authors also set up criteria for a *veridical* process tracing approach which are bound to the following restrictions: subjects should only report data from short term memory in oral form which guarantee not too complicated processes. Given these cases, the resulting process should be veridical.

3.2 Information boards and computerized methods

The second approach to be discussed in this chapter is *information boards*. With information boards it is possible to record the decision maker's information search without interfering with him (Payne, 1976). The term "board" has to be understood literally with the earlier versions of this method because the participant was asked to choose items from a matrix (alternatives X attributes) on a card*board*. To obtain information a card had to be pulled out of an envelope (cell) on the board. The amount and sequence of the cards looked at was written down and analyzed afterwards. This basic idea was already transferred to the computer in the pre-mouse era (Apple introduced the first mouse on a Macintosh in 1984, therefore broad public usage was not reached before the late 80s) when Payne and Braunstein (1978) or Dahlstrand and Montgomery (1984 as cited in Payne et al., 1993) used the keyboard as an input device. On a monochrome screen subjects saw the matrix (alternatives X attributes) with numbers in each cell for, e.g., different apartments to choose from. After entering a number of a cell on the keyboard the content of the appendant cell was displayed. With the entering of the next digit this cell was covered again and a new one was opened.

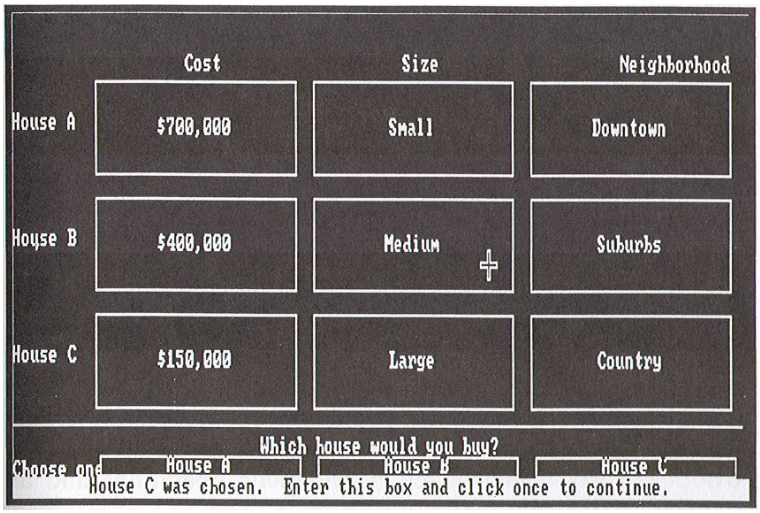


Fig. 1. Screenshot of original MouseLab version. Source: Payne, Bettman and Johnson, 1993, p. 275

More useable experiments were possible through the introduction of the mouse as an input device in psychological research. Here especially *MouseLab* (Payne et al., 1993) is a very popular tool. Participants see a matrix on the computer screen. Through moving the computer mouse over one of the boxes (in some versions additionally clicking on the box was necessary) the information underneath is shown (see Figure 1 for a screenshot of the original MouseLab version). When clicking on the next information the previous one is covered again. This system made it possible to track the sequence of information accessed and to record the duration information was looked at. MouseLab served as the central experimental tool for development of the Adaptive Decision Maker Theory by Payne et al. (1993, see Chapter 4.3, p. 28 for details).

MouseLab is still used in information board research and has been further developed. Levin, Huneke, & Jasper, 2000 combined the *phased narrowing technique* with information boards, whereas Jasper & Shapiro, 2002 developed a Windows version of MouseLab called MouseTrace with additional features like multiple responses and decision stages.

A third approach introduced in this chapter – *eye-tracking* – can be regarded as the most elaborated method (in terms of technical requirements) of recording an information search process. Two systems are available for this purpose a *head-mounted* and a *remote-mounted* one. In the head-mounted approach participants have to wear a helmet on which the camera for recording the eye movements is installed. In a remote-mounted system a less inversive recording is possible through using a camera which is mounted on the table in front of the subject or on a computer monitor (Lüer, Lass, & Shallo-Hoffmann, 1988). Lohse and Johnson (1996) compared MouseLab with an eye-tracking system called Eyegaze. Fixations (a fixation defines a threshold for the minimum resting time of the eyes in an area) were recorded while solving the same decision board task with Eyegaze and MouseLab. The usage of an eyetracking tool resulted in less time use, more fixations (a fixation in the MouseLab system was the stopping with the mouse in an area), less total information search and more variability in the search process. The authors concluded that the more sophisticated system (eye-tracking) reveal slightly different results, but basically deliver the same messages as MouseLab does.

3.3 Online research and process tracing

With the introduction of the Internet not only a change in communication behavior was elicited but also a possibility for a new research area born. Online research (see Musch & Reips, 2000 for an overview of the method) grew over the last years and presented results in such diverse fields as: questionnaire research, experiments, social interactions or cross cultural studies (Birnbaum, 2000; Reips & Bosnjak, 2001; Batinic et al., 2002).

Already at an early stage of online research Lohse and Johnson (1996) stated that the usage of Internet based tools (e.g., clickstream analysis of consumer decision making) should bring an important improvement for future research. Additionally, they suggested changing the information board approach (which has limitations in usability when it comes to large numbers of alternatives or dimensions) to a more naturalistic way of research. This aim should be accomplished by introducing a keyword-search methods (entering of a keyword and browsing through results) which is used with common search engines in everyday life.

Further traces of process tracing techniques using the online method can be found in consumer research (Johnson, 2001) or information sciences (White & Iivonen, 2001; Loeber & Cristea, 2003). Experimental research in decision making was done by Fasolo, McClelland, and Lange (in press) who used a Web-based decision board called WebIDB (Web-Information Display Board, based on the MouseLab idea) in a Web-experiment. The authors manipulated attribute characteristics of consumer goods (cameras) and inspected aids and strategies for decision making. With positive correlations between attribute unaided choice processes were attribute-based and corresponding choices easy. In this scenario non-compensatory sites performed better. With negative correlations

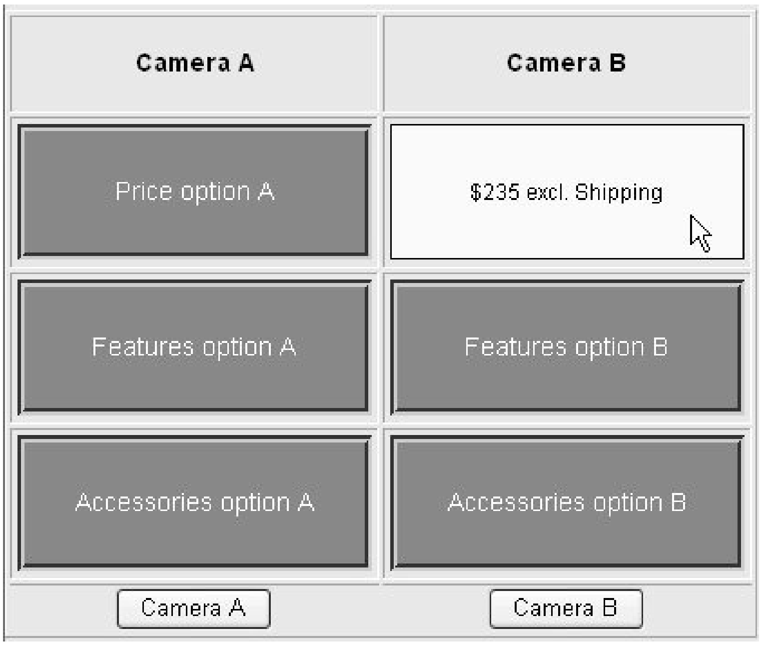


Fig. 2. Screenshot of MouseLabWeb example. Source: www.mouselabweb.org

Fig. 3. Screenshot of MouseLabWeb Structure. Source: www.mouselabweb.org

unaided choice processes were more alternative-based. The choices turned out to be difficult here and compensatory sites did better.

A recent development from the above described MouseLab software was introduced by Willemsen and Johnson (2004) who transferred the process tracing tool from the laboratory to the Web (see Figure 2 for a screenshot of a participant's view and Figure 3 for a screenshot of the settings screen, which provides insight into the various parameters, that can be defined within an experiment). In an Open Source project on http://www.mouselabweb.org the software can be downloaded and modified by anybody.

Further discussions about online research and decision making can be found in Chapter 6, p. 53.

3.4 Active information search

Antecessors

Engländer and Tyszka (1980) were among the first who addressed the issue of creating more realistic process tracing tasks in order to overcome the artificialness of decision boards. The authors criticized the prestructuring in decision boards where attributes and alternatives were labeled for the participants.

The experimental setup in Engländer and Tyszka (1980) included the participant, an official experimenter and three experts (collaborators of the experimenter) who answered questions. Answers were prepared before the experiment and balanced to present equally attractive alternatives. Participants were given two decision problems (a job selection decision – three jobs to choose from and an editorial decision problem – which book to publish). During the experimental session participants were allowed to ask the "experts" as many questions as they needed to be able to make a decision.

Results showed that attributes were investigated six times on average. This rather small number was surprising given the fact that much higher numbers were investigated with decision board tasks. Engländer and Tyszka (1980) explained this finding with the open structure of the decision task, participants may have simply missed some alternatives and therefore the average number of questions went down. With the calculation of a thoroughness index the authors found a larger interest (higher thoroughness) for the actually chosen alternative than for the one not chosen. This result is in line with SDS and DiffCon theory and can also be found in many studies by Huber and colleagues (see below).

The method of active information search

The demand of Engländer and Tyszka (1980) for the necessity of a new research paradigm in view of shortcomings in traditional decision making research can be found in Huber et al. (1997) again. Especially the fact that a vast majority of decision making research was done with gambling tasks or tasks which were very abstract drove the authors to develop something new. Huber et al. (1997) called their new method *Active Information Search (AIS)*. AIS should guarantee uninfluenced gathering of information in a decision making task. Furthermore within the application of this method tasks should be situated in a more naturalistic or quasi-realistic setting without the focus on experts as suggested in Klein, Orasanu, Calderwood, and Zsambok (1995) or Zsambok and Klein (1997).

Experiments with AIS have the following structure: Participants get a short description of the decision situation and the possible alternatives to choose from. Then the participant is allowed to ask (any and an arbitrary number of) questions about the task and gets answers which have been prepared and printed on cards before the experiment. The already selected cards remain on the table and are often reorganized by the participant in order to represent alternatives, attributes or other criteria developed (see Figure 4 for an example of an experimental setup). As soon as the participant is satisfied with the gathered information he states a choice between the alternatives.

During the gathering process the experimenter records the collected information (the sequence of gathering).

Fig. 4. Instructor (left) with a participant in an AIS session

Huber (1997) categorized the information presented to the participants into several classes. Unfortunately, these classes changed over the years, which makes them not directly comparable. Nevertheless the following categories were used in Huber et al. (1997) as well as in Huber, Beutter, Montoya, and Huber (2001): probability, new alternative, control and worst-case plans. The usage of *probability* information was relatively low (or zero) in tasks from the domain of management and economy. More usage of probabilities was found in a task from the medical domain. *New alternative* differed between technical (small usage) and the other domains (managerial, economic, medical) where medium usage could be found. *Control* was equally low in technical and managerial tasks when compared to the other two domains. Finally, worst-case plans were only of interest in the medical and managerial domain but not in the other two domains.

The method of AIS has been developed further over the following years and resulted in various papers dealing with a range of experimental conditions. Whether probabilities were recalled from background knowledge was addressed in Huber and Macho (2001). The authors showed that in quasi-naturalistic tasks participants did not infer probability information from background knowledge. Huber and Huber (submitted) addressed the differences between gambles and other task types. The authors showed that with a variation in cue quality (rich versus impoverished cues), it was possible to change the interest in certain information types. To guarantee no influence of answers onto one of the following questions (which is inherent to the AIS method) the questioning and answering parts were separated. All questions of interest had to be asked first and answers were given all in one afterwards. By this separation no influence was guaranteed. The amount of information of interest was approximately the same as in the standard AIS version which provides support for using this technique. Results showed that the richness of the cue influenced the amount of information wanted, too. The *richer* the cue the more information was demanded. The type of information demanded also varied with the type of task: probability information was prominent in gambles whereas information on risk defusing operators (see p. 18) was rarely demanded. A reversed pattern was found in the other tasks – showing that there is a very basal difference between the gambling and the other tasks.

The AIS method was not only tested in a laboratory situation, two studies used this method in external settings – Schulte-Mecklenbeck and Huber (2003) applied the basic principles of AIS in a Web study (see Chapter 6, p. 53 for details) whereas Shilo, Gerad, and Goldman (under review) used the method in a genetic counseling task. In this real life situation the most important issues for participants were outcomes and consequences of the given alternatives. Following this, possible risk elimination methods (see Risk Defusing Operators below) and background knowledge were of interest. Access to probability information was somewhat awkward because participants only wanted "definite information stated with certainty" (Shilo et al., under review, p. 16). From a more theoretical perspective one special class of information called Risk Defusing Operators (RDOs) will be addressed next.

Risk defusing operators

Huber (2004), p. 7 defines RDOs as ...

> "... an action intended by the decision-maker to be performed additionally to a specific alternative and [as being] expected to decrease the risk."

Certain types of RDOs can be distinguished where (1) *the time* when the RDO is applied and (2) *the target* of the RDO is of central interest. In respect of time *pre-* and *post-event* RDOs are possible. An example for a pre-event RDO is the backing up of computer files in order to be prepared for the case of a system crash. Whereas a post-event RDO could be medical treatment after a disease has started. The post-event RDO is clearly superior to the pre-event RDO in most cases because action is only required *after* a negative event took place. On the second dimension – the target of the RDO – an *outcome* can either be *prevented*, e.g., with a vaccination against a disease or an *outcome* can be *compensated* with, e.g., the buying of insurance (against loss of income) when a credit is taken.

Several factors may influence the search and the acceptance of an RDO. The following are described in more detail: in terms of searching: *attractiveness of an alternative* and *expectation of search success*; in terms of acceptance: *Cost, Effect* and *Success Probability*:

(1) Huber et al. (2001) showed that search for an RDO is mainly performed for a *promising alternative* (compare to Chapter 4.1 and Chapter 4.2, p. 21 et seqq.). This is in contradiction to e.g., the phases of decision making postulated by Kahneman and Tversky (1979) where an elaboration follows a representation phase. Huber (2004) suggests a mixture of these two phases where evaluation and construction are overlapping parts.

(2) *Expectation of search success* should influence the basic demand for RDOs. The higher the subjective expectation of a successful search is the higher is the probability that a RDO will be searched for (Huber & Huber, submitted). Here domain specific knowledge plays an important role. If the decision maker is familiar with the decision domain a judgement of the probability of an RDO in the domain is possible. In an unfamiliar decision domain this judgment is harder to make and therefore information demand should be orientated to other information categories.

(3) *Costs* of applying an RDO and related side effects (which could result in new costs) should additionally be important to the decision maker. Williamson, Ranyard, and Cuthbert (2000) showed that in an insurance situation decision makers not only search for the RDO but are also concerned with costs the RDO may result in. An RDO was chosen more often when it was perceived as being cheaper.

(4) *Effect* of an RDO is described in terms of a decrease of the probability of the negative outcome. An example is a vaccination against an infectious disease, which may prevent the outbreak

of the disease completely or only helps overcoming it (Huber, 2004). If the negative event can be prevented with a high probability the choosing of this RDO should be frequent. A factor closely connected is the *success probability* of the RDO. It determines the probability of the intended effect of an RDO. In the vaccination example, the drug may not be successful for a minority of people who have been vaccinated. Again, domain knowledge plays an important role in the judgment of this probability. In absence of hints at the effectiveness of an RDO background knowledge can serve as a factor for judgment.

3.5 Summary of methods and applications

Summarizing the introduced process tracing methods, first of all a basic assumptions about how to approach decision making research, seems important. The *structural* approach is useful for observation of the input and output of a decision. The limitation of this approach is the ignorance of the process within the decision maker in most of the current research settings. This drawback can be overcome with the *process approach* which inspects processes happening between the input and the output.

Strategies that resulted from process tracing research can be divided into *compensatory* and *non-compensatory* ones (further discriminations will be introduced in the following chapter). A strategy that can outweigh one attribute to another is called *compensatory.* If outweighing is not possible a *non-compensatory* (easier) strategy is used. Often a combination of these two strategies, e.g., one of them following the other, can be observed.

Four methods (introspection is not seen as an up to date method in this overview) were introduced in this chapter. Each of them fits into a specific research setting, but also the combination of certain methods for gaining more insight seems promising. *Thinking aloud* is a good example for a "combination" method. Studies with information boards, eye-tracking or AIS were all run in combination with thinking aloud protocols. The thinking aloud data can either be used to strengthen an effect found with other methods or disclose effects which are not accessible otherwise. This type of data also provides direct insight into what the participant is actually *doing* in a task. For example: Does a participant multiply outcomes and probabilities after looking up the two information items in an information board?

Using *Information boards* as a research method bears the advantage of a controlled settings in which the participant searches for information. The biggest drawback of the method clearly is the problem of prestructuring. Most often information boards follow an alternative X attribute structure with ready named dimensions. This prestructuring could lead to other results (participants might look up information they wouldn't have thought of otherwise) than with methods that provide no such structure. The combination of information board studies with, e.g., eye-tracking revealed more insight into the time *between* actual behavioral actions of the participant (picking up of a card or clicking with a mouse).

Eye-tracking research is closely related to information boards, however the possible applications concerning the used stimuli, is much broader. A participant can basically interact with anything displayed on a computer screen. Certain boundaries of information boards like the reduction to alternatives and attributes do not apply within this methods. Eye-tracking research is widely used in such diverse areas like, e.g., army pilot training (Cheung & Hofer, 2003) or analysis of handball in sports (Pers, Bon, Kovacic, Sibila, & Dezman, 2002).

The fourth method discussed in this chapter – *AIS* – was introduced as an advancement of standard decision making tasks like gambles. In these standard tasks all necessary information is presented to the participant, which is no longer the case in AIS where only information is provided which is actually needed by the participant. This is done without labeling of alternatives or prestructuring of a decision situation (as, e.g., in decision boards). Thus a combination of AIS with, e.g., eye-tracking seems reasonable, no such study was done up to now.

In this overview *thinking aloud* clearly dominates in terms of compatibility. It sheds light on the actual actions going on within the decision maker which are not directly observable with the other introduced process tracing methods. Despite this benefit the other methods show consistent results between studies, too. A reason for considering and/or combining all (or at least some) of them.

4

Theoretical approaches

The following theories were identified as being relevant for this thesis: H. Montgomery's *Search for Dominance Structure* (1989) and Svenson's *Differentiation and Consolidation Theory* (1992, 1999). Their predictions of certain stages in the decision making process and their emphasis of a process tracing approach fit to this thesis' intended research questions. Furthermore *The Adaptive Decision Maker* (Payne et al., 1993) approach was selected because of (1) the methodological relatedness of MouseLab to the methods used here, (2) the explicit mentioning of the "domain differences" in the theoretical outline and (3) the demanded adaptiveness of participants to certain task and domain situations.

4.1 Search for dominance structure (SDS)

H. Montgomery (1989) as well as H. Montgomery and Willen (1999) describe a decision as a preparation for action. This *intention to act* has to be *defended* and *supported* by the decision maker. Therefore, it is important for the decision maker to be able to find good reasons for his choices. Being able to identify a *dominance structure* is recognized one of the ways to find a particularly strong justification for a choice. The case of one alternative dominating the other gives the decision maker optimal possibilities for defense against others or himself. The outlined theory concludes that the decision maker will be motivated to find a dominating structure. If no such structure exists he will restructure the problem in order to generate such a structure.

H. Montgomery and Svenson (1989a) suggest that the decision maker goes through four stages to achieve a dominance structure (see Figure 5, p. 22): (1) in the *pre-editing phase* the decision maker identifies alternatives and attributes relevant to the decision process. Attributes which are not regarded as important and alternatives with small chances of becoming dominant are discarded. Already at this early stage, alternatives which are very unattractive are removed from the decision process.(2) *Finding a promising alternative* is regarded as the second step in which the decision maker forms a preference for one alternative that is temporarily regarded as more attractive than other alternatives on one important attribute. (3) The *dominance-testing* phase checks whether the identified alternative is superior to the other alternatives (there should exist no disadvantages and at least one advantage). If this is the case the alternative is chosen. If not, a check is performed whether all relevant pieces of information has been taken into account and dominance testing starts all over again. More frequent however is the proceeding to the fourth and last phase. (4) In the *dominance-structuring* phase a restructuring of the situation (information) is carried out. Several operations can be performed in this stage: A *de-emphasizing* operation is used to reduce the importance of a disadvantage or of a difference between alternatives concerning the promising option. A *bolstering* operation increases the advantages of the promising option or

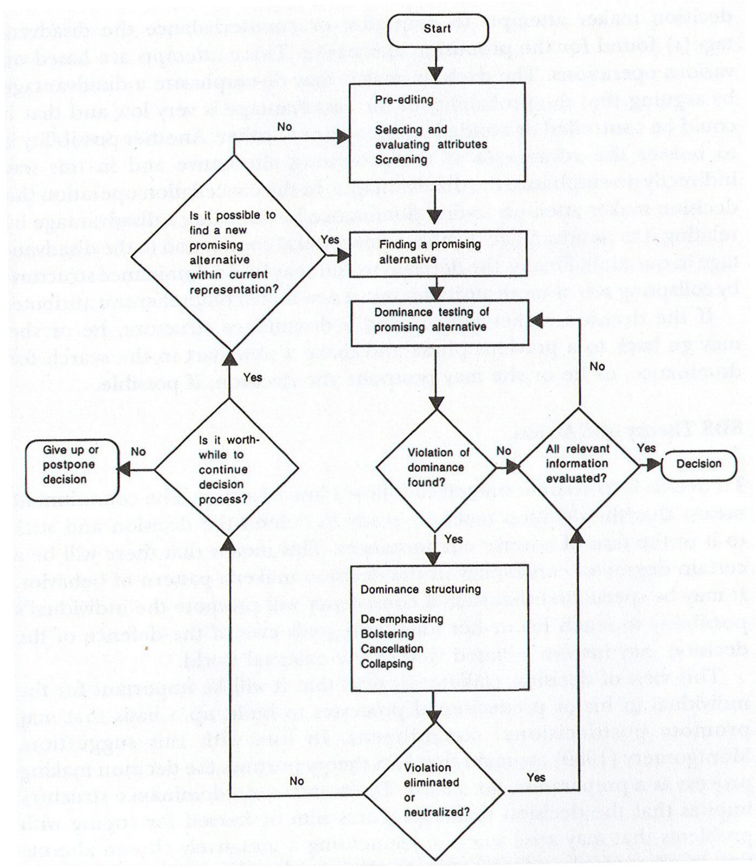


Fig. 5. Dominance Search Model. Source: Montgomery & Svenson, 1989, p. 25

decreases those of the others. H. Montgomery (1989) argues that both methods are a "somewhat irrational" way of finding a dominant alternative. The author suggests two further – more rational – actions: A *cancelation* operation rules out a disadvantage through the association with a (logically connected) advantage. A *collapsing* operation ties together different attributes in a single dimension (for instance, using a time or money scale). If a dominant alternative can be identified through this process – it is chosen. If not the process starts again from the *pre-editing* phase or searches for a *new promising alternative*. The purpose of the restructuring process is to end up with a simple representation of the decision problem and the possibility of applying a simple decision rule – *choose the dominant alternative*. In doing so, the decision maker has not only the advantage of an easy (final) process for his decision but also the guarantee of easily being able to justify the decision to himself and others.

Studies

Two empirical sources (H. Montgomery & Svenson, 1989b; Backlund et al., 2003) demonstrating the predictions of SDS in experiments are discussed now. Methodological aspects of both studies were already mentioned in Chapter 3.1, p. 11. Here, only thereof resulting interpretations are of interest.

In H. Montgomery and Svenson (1989b)'s study the participants were confronted with five booklets containing information about five different houses. In individual sessions subjects were thinking aloud while working through the booklets. The sessions were recorded and coded by means of a coding schema containing attributes like economic aspects, distance from city, ground plot of house or services in the area. Protocols were divided into four parts to be able to separately analyze changes in attention to certain attributes. Two different measures were coded: (1) *attention* denoted the number of referrals to an alternative, whereas (2) *evaluation* was defined through a ratio of positive and negative statements relative to all statements.

The authors found a reversal of attention during the decision process. In the first phase (at the beginning) the finally not-chosen alternative received the most attention. In the fourth phase (at the end of the process) this pattern was switched as now the chosen alternative received a magnitude of attention. For evaluation a not very surprising picture emerged with a more positive evaluation of the later chosen alternative in comparison with the not chosen one. These two patterns together are interpreted as reflecting the process of choosing an alternative and realizing during *dominance-testing* that it is not the best available one. Through further analysis and evaluation of attractiveness, evidence for a *dominance-structuring* phase was found.

However a critical remark about this study and its interpretation has to be made. H. Montgomery and Svenson (1989b) had only a very small sample of 12 participants. This situation results in even smaller samples of no more than six per group, when the authors split their sample for some analyses (e.g., to check if there is different behavior in participants who turn to the later chosen alternative at the beginning in comparison to those who switch preferences during the decision process). Despite the superficially clear cut message, this fact lessens explanatory power of the study.

Backlund et al. (2003)'s paper is not only of methodological interest (the question whether rating scales could do as well as thinking-aloud procedures were raised) the study also tested the predictions of the SDS theory in connection with a medical decision making task. The authors predicted an increasing interest of participants to the finally chosen alternative in the course of information search. As already mentioned above, information was presented on a computer screen. On each screen, one group of participants had to rate (rating) their will to prescribe a drug or not while thinking aloud (providing a statement). The second group did thinking aloud (without the ratings) during the task.

When comparing the ratings between the two groups over time (the consecutive screens) in five of the six tasks, a differentiation in confidence ratings between the "prescribe drug" and "do not prescribe drug" deciders was clearly found. This result speaks for a strong differentiation phase which already starts off early (with the second decision in this study). For the thinking aloud group results were not that clear. Again, one case (the same as in the ratings group) resulted in no differentiation because clearly all decisions were "not to prescribe" a drug. In the other five cases, differentiation (an increased distance between the direction of statements) was found, but much later (around the fourth decision) in the process. In terms of the question whether the addition of a rating scale influenced the outcome – no differences were found concerning the final decision between the *think-aloud* and the *think-aloud + rating* group. Despite the little more time that was needed by the *think-aloud + rating* for completion of the tasks (the mean differences between groups were around one minute), an inspection of the number of statements revealed no differences between the two groups. Backlund et al. (2003) concluded that verbal protocols are at least as sensitive at revealing a differentiation between alternatives in a process tracing study but without providing insight into, e.g., strategies of the decision maker, which solely can be found with the thinking aloud method.

4.2 Differentiation and consolidation (DiffCon)

Before DiffCon theory is outlined, some general remarks from Svenson (1990) on classifications of decision situations are made. Svenson differentiates two very basic *types* of decision situations which he calls *type A* and *type B*. *Type A* decisions have two alternatives where one of them has to be chosen. In contrast, *type B* decisions have a status quo or an option to develop new alternatives. Type B decisions might therefore arise before type A decisions (see Figure 4.2, p. 24). In decision making research type A decisions are common – alternatives are normally presented to subjects to choose from. The development of new alternatives, a situation happening quite regularly in real life, is demanded seldomly in an experimental setting.

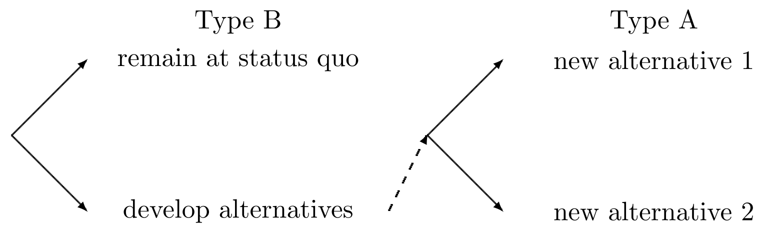


Fig. 6. Svenson's Decision Types A and B. Adapted from: Svenson, 1990, p. 18

Additional to the type A and B differentiation, a central concept in Svenson (1990)'s approach (see Figure 7) is that a decision is based on the decision maker's *values* which are mapped onto the current *decision problem*. In this process, *four levels* of "depth of processing" are distinguished. The processes on level one (1) are recognizing a decision as being *already performed in the past* and choosing the alternative that is similar to the already chosen one (compare to Maule & Pearman, 1994's rule based category). (2) On level two, an *evaluation of alternatives* is added. This evaluation can contain emotional or motivational processes (compare to Arroba, 1977's emotional category). Decision rules are non-compensatory, i.e., a bad attribute on one alternative can not be compensated by another attribute or alternative. (3) On level three, decision makers actively build connections between their value system (see the outer right part of Figure 7) and the decision alternatives. A conscious process of weighing alternatives against values takes place. (4) Finally, on level four, the decision maker has to create new alternatives in an unfamiliar decision situation.

The already mentioned type A and B decisions can be incorporated into the four level model with the following hypothetical decision situation. The decision maker recognizes a decision problem and checks whether it is a type A (has two alternatives?) or type B decision. In case of type B, a *planning* stage determines the level and the goal of the decision as well as a strategy for reaching this goal. A subsequent *information search* helps creating alternatives (in case of a new decision problem) or goes back to a modification of planning. In the *creation of alternatives* stage, the later very important identification of a *promising alternative* takes place. After a check of the available alternatives and the elimination of not-acceptable alternatives, a comparison to the before defined promising alternative takes place. In case of dominance, the alternative is chosen and certain post-decisional processes – mainly concerning the attractiveness of the alternatives – take place. In case of no dominant alternative having been found, the process starts again either until a dominant or until only one alternative is left (which should at least dominate the second one).

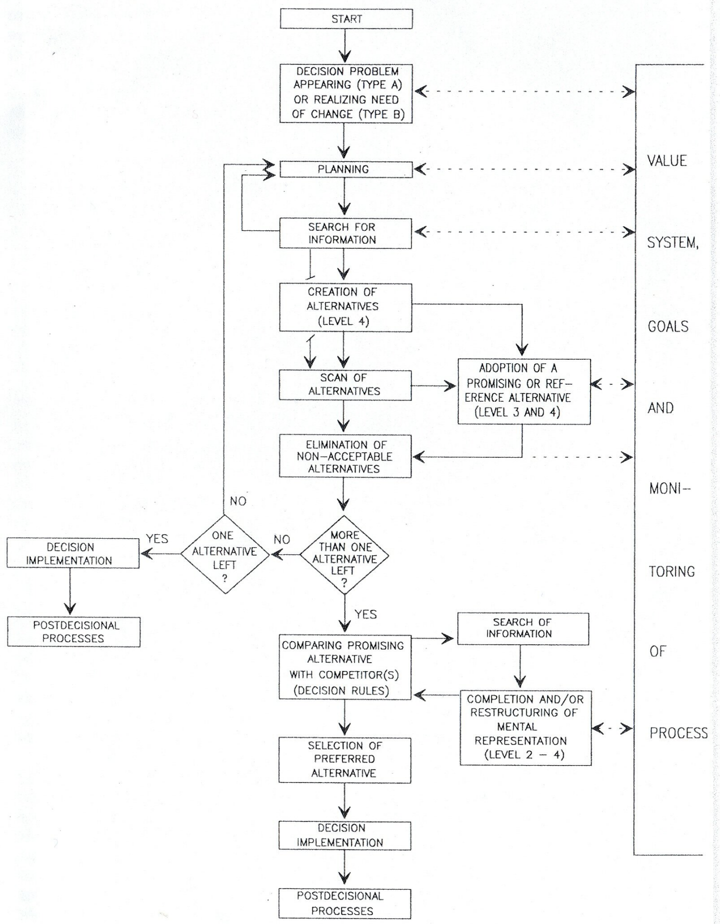


Fig. 7. Svenson's Process model. Source: Svenson, 1990, p. 24

The Theory

More recently, Svenson's model was further elaborated and named *Differentiation and Consolidation Theory* (DiffCon) (Svenson, 1992, 1999). A lot of the concepts from the 1990s' model can be found again in these two papers. The decision making process is initially influenced by *markers* (e.g., the label of an attribute – cost) which help to make one attribute the most important one. A marker can trigger a very rapid stopping rule like "this is a very cheap price – buy it!". The choice of a reference or preliminary choice alternative helps setting a reference point to which all other available alternatives can be compared.

The comparison of the preliminary alternative to the rest of available alternatives is called *differentiation*. The aim of differentiation is to always create a sufficiently different alternative which is then chosen. Three types of differentiations are suggested by Svenson (1999): *holistic*, *process* and *structural* differentiation. (1) Holistic differentiation is a quick and *low level* method which is often regarded as a classification process. Emotions are an example for quickly bringing the decision maker to results without extensive, deliberate considerations. (2) The usage of *decision rules* (compare to Chapter 4.3, p. 28) characterizes the stage of *process differentiation*. It "depends on the individual, the context and the structure of the decision problem" which rule is used (Svenson, 1999, p. 178). (3) *Structural differentiation* as a third type demands a change of very basic components of the decision process like attractiveness of alternatives, importance of attributes or structure of the decision problem per se. After the decision, the process is not over but a *consolidation* of the chosen alternative's advantages in comparison to the not chosen alternatives takes place. This is basically the same process as before the decision but it helps the decision maker to overcome potential threats which are caused by the decision outcome.

Several predictions are derived from this process model. It should be possible to identify the chosen alternative through the greater degree of attention it receives in the differentiation process before the actual decision is made. Further, and more important, real, irreversible decisions rich of consequences should be differentiated more than a reversed type. If this assumption is true it holds a possible explanation for different information gathering and decision making behaviors when different domains of decision making are involved.

Studies

The conceptual idea of DiffCon theory was applied in various research designs. Two exemplary studies are discussed now. Svenson & Malmsten (1996, as cited in Svenson, 1999) studied children who had to decide between getting tickets for two lotteries with headsets as prices. The first lottery had a high quality headset as a prize with a smaller chance ($p = \frac{1}{3}$) of getting a ticket. The second lottery had a low quality headset as a prize with a higher chance ($p = \frac{1}{2}$) for the ticket. Participants chose one of the lotteries, then it was determined whether or not the chosen lottery would be played.

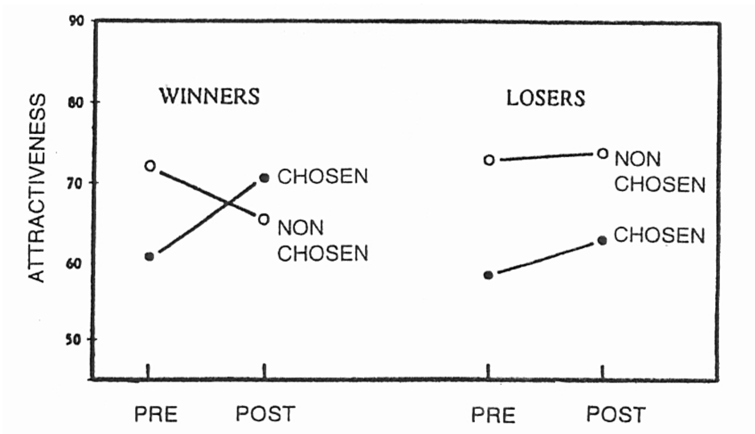


Fig. 8. Changes in attractiveness ratings between pre- and post decision. Source: Svenson, 1999, p. 192

Figure 8 depicts the results for the "winners" and "losers" of a lottery ticket. In the winning group there is a switch between the pre- and post-decision attractiveness rating, i.e., an increase of attractiveness for the chosen and a decrease in attractiveness for the not chosen item was found. Contrary to this result, the "losers" group clearly rated the "not chosen" item as more attractive in the pre- as well as in the post-decision. Svenson & Malmsten (1996, as cited in Svenson, 1999) emphasize that the restructuring process that was initiated in the "winner" condition through knowing the outcome was so strong, that attractiveness ratings were reversed. Svenson et al. tried to replicate this result with more realistic outcomes through testing students and their decisions about courses to take in the following year. In this setting, the projected switch was not as strong as in the first study (again using attractiveness ratings). The authors drew the conclusion that attractiveness rating might not be an appropriate measure for a decision between courses, because a less attractive course might be chosen because of other reasons (importance for future jobs ...).

Feldman-Stewart, Brundage, Manen, and Svenson (2004) used DiffCon theory in a study on decision aids for patients with prostate cancer. Participants were interviewed before (three times) and after (two times) a treatment decision between (1) surgery, (2) radiotherapy or (3) a no treatment option. In each stage a rating on *treatment preference* (TPA) was collected from the patients. This preference contained a rating of all alternatives on a five point scale.

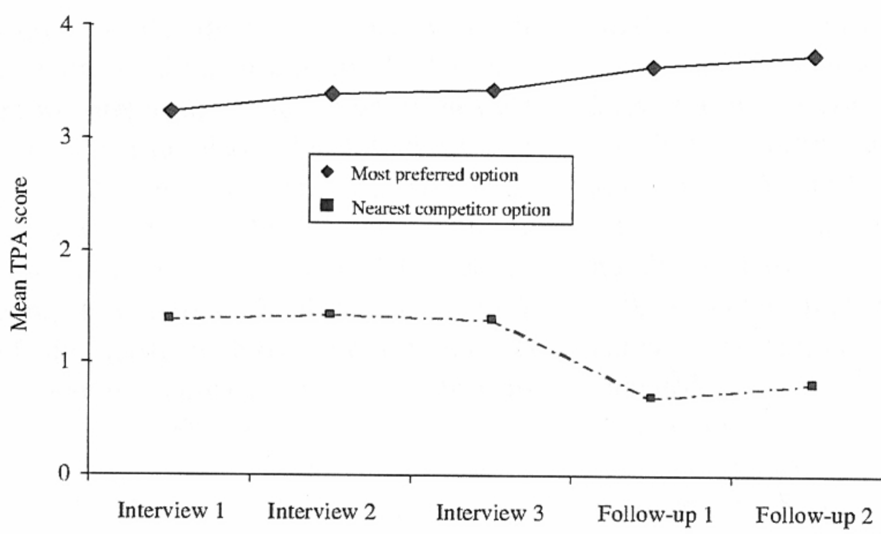


Fig. 9. Differentiation process between two cure alternatives. TPA – treatment preference assessment. Source: Feldman-Stewart, et al., 2004, p. 138

Results indicate that between patients there is a large variation concerning attributes important for a decision. Furthermore the number of attributes considered as relevant to the decision was huge (This result is in sharp contrast to classical decision making experiments where only a limited number of information is presented to the participants.) A central aim of the study was to demonstrate a *differentiation* and *consolidation* process demanded by DiffCon with “real life data”. Figure 9 plots the five interview times and the mean TPA scores for the most preferred option and its nearest “competitor”. A stable preference with a small trend of increasing attractiveness at the follow-ups (after the decision) was found for the favorite alternative. Contrary to this, an at first stable judgment of preferences for the second best option decreases significantly at the two follow up dates. The differences between the most preferred alternative and the follow-up in the interviews reflects the predicted differentiation phase. The consolidation phase is even more accentuated when the two alternatives disperse in the two follow up conditions. Feldman-Stewart et al. (2004) conclude that there is a large variation in the relevance judgments of attributes between patients. This fact has immediate consequences when the generation of a decision aid is of interest.

The authors suggest that, e.g, the number of attributes in an aiding process should cover a wide range of interests to be able to fit to as many patients as possible[1].

4.3 The adaptive decision maker (ADM)

As a third theoretical approach the ADM is discussed. Payne et al. (1993) describe a comprehensive theory for how decision makers adopt to a given decision situation. The final decision of "how to decide" is either a conscious choice or a learned contingency between *elements of the task*, the *effort* and the *accuracy* of a decision strategy. The authors see decision making as a "highly contingent form of information processing" (Payne et al., 1993, p. 68) and emphasize the importance of finding principles that describe information processing for a given task in a general way. These principles are formulated as the following heuristics:

Table 1. Properties of choice heuristics.

Heuristic	Compensatory versus non-compensatory	information ignored?	attribute- versus alternative-based
WADD	C	N	AL
EQW	C	Y	AL
MCD	C	Y	AT
ADDIF	C	N	AT
FRQ	C	Y	AL
EBA	N	Y	AT
SAT	N	Y	AL
LEX	N	Y	AT

- The *Weighted ADDitive rule* (WADD) (see Table 1, adapted from: Payne, Bettman & Johnson, 1993, p. 32) takes the values of each alternative on all relevant attributes in account. Furthermore, it considers all weights the decision maker ascribes to the attributes. WADD is often considered the *normative* standard for decisions. It requires intensive processing because for each alternative the weight times the value for each attribute is considered. After the weighted attributes are summed up the alternative with the highest value should be chosen. Two popular strategies which are related to the WADD rule are the calculation of an expected value (EV) of an expected utility (EU). EV is calculated in gambles as the product of the outcome times the probability of occurrence. When exchanging the value of an outcome with its utility, the EU rule is also applicable to non-gamble situations. However, the intense usage of calculations makes WADD somewhat unlikely for usage in a real life setting. A common assumption in the ADM framework is the usage of simplifying heuristics to ease the processing of data (in a recent approach with *fast and frugal* heuristics, this idea is extensively used – see, e.g., Gigerenzer et al., 1999 or Gigerenzer & Selten, 2001 for an overview).
- The *EQual Weights* (EQW) heuristic simplifies the WADD rule in ignoring probability or the attributes' relative importance. It calculates a sum of values for each attribute on an alternative. As a pre-condition, it is necessary to be able to express values for the attributes on a common scale.

[1] This results can also be found in the WebDiP study p. 77 where participants interest in information ranged over the whole spectrum of available information. The requested information is considerably larger than the necessary information when the structure of the tasks is considered.

- The *Majority of Conflicting Dimensions* (MCD) rule is a special case of the more general *ADDitive DIFferences* (ADDIF) rule. The MCD rule simply compares pairs of alternatives on all given attributes. The alternative with the better attributes is chosen. ADDIF rule compares two alternatives directly on each attribute and determines a *winner* of the two alternatives through summing up the difference scores. The winner is compared to the next alternative of interest with the same procedure. At the end of the process, an overall winner should result (this procedure gives the same winner like in a MAUT analysis with a different calculation process!).
- The *Frequency of good and bad features* (FRQ) rule counts good and bad features of alternatives. It is necessary for the decision maker to develop a cutoff point for a "good" and a "bad" feature. Depending on the focus of the decision maker, different versions of this heuristics are created.

Turning from compensatory to non-compensatory decision heuristics, the following three are presented in Payne et al. (1993):

- *Elimination by aspect* (EBA) rule consists of two steps: (1) The most important attribute is determined. (2) The cutoff value for this attribute is retrieved and all alternatives below it are eliminated. This process is rerun with the second-, third-important ... attribute until only one alternative remains.
- The *Satisficing* heuristic (SAT) considers alternatives one after the other in the order they occur. Each attribute of an alternative is compared to a cutoff-level, if an attribute is below this level the alternative is rejected. The *first* alternative that meets all requirements (is above all cutoff levels) is chosen. An interesting point in this heuristic is the emphasis on the order of alternatives in a set. Having a good place in the order enhances the possibility of being chosen and vice versa.
- The *LEXicographic* (LEX) heuristic chooses the most important attribute and evaluates all alternatives on this attributes. The alternative with the best value on the most important attribute is then chosen. If there is a tie, the second most important attribute is taken into consideration.

Two studies are described now which test the above introduced heuristics for performance under time-pressure (Rieskamp & Hoffrage, 1999) and from the perspective of the decision maker's performance perceptibility (Chu & Spires, 2003).

Heuristics, predictions and time pressure

The performance of choice heuristics under time pressure is explored in Rieskamp and Hoffrage (1999). The authors use EBA and LEX which have already been introduced above. Additionally, the *Weighted Pros* heuristic (Huber, 1979) and variations of LEX (i.e., LEX-Semi (Luce, 1956, as cited in Rieskamp & Hoffrage, 1999, p. 143) and LEX-ADD) are inspected. A shift from compensatory to non-compensatory (from extensive to simple/frugal) strategies is predicted in the event of increasing time pressure. Additionally, more information about the most important cue should be acquired when time pressure is high (concentrate on the important information to solve the problem within the time constraints).

In an information board study using MouseLab (see Chapter 3.2, p. 13) participants had to decide between companies in terms of highest yearly profit. In the *low time pressure* condition *50s* (seconds) were available for information search whereas in the *high time pressure* condition only *20s* were available to scan though a matrix of 6 attributes of 4 companies (24 cells). Rieskamp and Hoffrage (1999) compare two methods of analysis, a process analysis and an outcome analysis, to identify the best heuristic.

The following results were obtained: (1) a attribute-wise search was especially common under high time pressure. In this condition also the average viewing time of an information item was lower

(325 ms) in comparison to that measured in the low time pressure condition (430ms). However this result was predicted by many heuristics and therefore a differentiation was not possible without including the resulting choices. (2) When the outcomes of the decisions were included it was possible to predict participants behavior best: under low time pressure with the Weighted Pros rule; under high time pressure with the LEX rule. (3) The authors also found an interesting connection between the time spent on a cue and its validity. Especially under the high time pressure condition, the correlation between time spent and validity was high. Participants were able to focus their attention on important cues ignoring the other information.

Perception of accuracy

From the perspective of how accurately a decision maker behaves in a task, the *perceptibility of accuracy and effort* in decision making situations was addressed by Chu and Spires (2003). The authors did not follow, e.g., Payne et al. (1993)'s strategy of deductively deriving strategies either from simulations or analyses of elementary information processes (EIPs). Chu and Spires's approach included a presentation (with a learning phase) of different decision making strategies (e.g., WADD, EQW, MCD, EBA or SAT) and a subsequent questionnaire dealing with each of these strategies. The aim was to demonstrate that basic assumptions about accuracy and effort of these strategies should be perceived by the participants in a way comparable to the theoretical assumptions.

Results indicate that WADD rule is perceived as highly accurate and very effective. On the other end of the continuum, a *random strategy* (RAN) had the lowest accuracy ratings combined with the lowest effort ratings. A second important result was that *intra*attribute processing (like in EBA, LEX and MCD) is perceived as being less effortful than *inter*attribute processing (like in EQW or WADD). On the dimension of compensatory versus non-compensatory strategies, the compensatory strategies (WADD, EQW) were perceived as being more effortful than the non-compensatory ones. The authors concluded that the perception (of effort and accuracy of decision strategies) of participants in a study is comparable to assumptions decision theorists have about these heuristics.

After the introduction of very basic assumptions about a decision process and introduction of common choice heuristics, two important effects concerning the decision task and the decision content are described next.

4.3.1 Task and content effects

Returning to Payne et al. (1993)'s framework, a decision process is decomposed into elementary information processes (EIPs). Using these EIPs, a decision making strategy (a sequence of mental operations to transform a state of knowledge into a final goal) is chosen based on a compromise between accuracy and effort required. This view of decision making is much in the light of Simon's (1978, as cited in Payne et al., 1993) ideas on *bounded rationality* (see also Gigerenzer & Selten, 2001 for recent usage of this concept). A consideration between cost and benefit takes place every time a decision is performed and depends among others on *task effects* and *context effects.*

Task effects include structural characteristics of the decision problem, number of alternatives, attributes or outcomes, time pressure and display modes. They are discussed according to the number of alternatives used (Payne, 1976, as cited in Payne et al., 1993), the response mode (Tversky & Kahneman, 1988, as cited in Payne et al., 1993) or the type of information display (Aschenbrenner, 1978, as cited in Payne et al., 1993). *Context effects* differ in respect to the values of objects, the similarity and overall attractiveness of alternatives. These effects are addressed with references to the framing effect (Tversky & Kahneman, 1981). Further connections are drawn to the similarity of alternatives (Luce, 1959, as cited in Payne et al., 1993) or the quality of an option set (Williams, 1966, as cited in Payne et al., 1993). Task and context effects will now be explained in more detail.

Task effects

The following task effects are discussed now: *task complexity, response mode* and *information display.* The relation of each of them to AIS and/or experiment 4 (Chapter 8, p. 77) will additionally be addressed (if applicable).

(1) Three parameters define the *complexity* of a task: the number of alternatives, the number of attributes and time pressure. With an increased number of alternatives a move from alternative-wise to attribute-wise information search can be observed. Whereas an increase in number of attributes additionally strengthens the subjects' confidence in their choices. More generally, the increase in complexity elicits a move from compensatory to non-compensatory strategies and the usage of simplifying strategies (heuristics). On the time pressure dimension, results are not that clear cut. Participants' reactions in studies vary from accelerated processing of information to perceptual narrowing of considered options or simple avoidance of a choice. Narrowing down alternatives to the most important ones is in line with Rieskamp and Hoffrage (1999) who analyze different decision strategies under low and high time pressure. Participants turned to simple heuristics (take the best) when time pressure was high (see above).

AIS: Up to now there are always two alternatives to choose from with many attributes in the AIS paradigm (see Chapter 3.4, p. 16). Because one of the aims of AIS is the uninfluenced gathering of information, no hints on possible dimensions are given in most of the experiments. Additionally, no time pressure experiments have been done yet.

(2) The *effects of response* modes have a long history in decision making research. Examples for different modes are the *choice* between two options where outcomes and probabilities are given; the *matching* between alternatives where one outcome or probability is missing; the *bidding* for a given gamble in order to play it or the *rating* of such a gamble in order to identify its attractiveness. All of these modes are primarily studied with gambles and are very robust across different studies.

AIS: The response mode in the AIS experiments is a choice. As mentioned above in all of the studies, there are two alternatives to choose from. An advantage of using choices as responses is that this mode represents a familiar situation for participants. Matching, bidding and rating seem to be awkward in many real life situations.

(3) *Information display* concerns the presentation of information to the participant. Bettman and Kakkar (1977, as cited in Payne et al., 1993, p. 49) found evidence that it is possible to encourage subjects to search for information, e.g., in an alternative-wise way through changes in the presentation format (e.g., "prize presentations" used in a supermarket which are familiar to the participant). This results is an important hint at how to run information search studies. A variation of information concerning the position on the display is crucial for eliminating such effects. Another display factor discussed is the completeness of information, e.g., missing parts of information for one item. Here, a compensation for the missing information through inferences from the remaining attributes can be demonstrated (Ford & Schmid, 1987, as cited in Payne et al., 1993).

Experiment 4: The intention of the used presentation mode in experiment 4 is to minimize effects like these. Participants start off with a screen on which only the task text and the keyword field is displayed. The listing of the actual information is done without hints on importance or preferences.

Context effects

The second group of effects influencing the decision maker is based on context variables. One of the very well known effects is the *framing effect* first described by Tversky and Kahneman (1981). For more recent reviews see Kühberger (1998) or Levin, Schneider, and Gaeth (1998). Here the wording of formally the same problem in either a positive or a negative way changes choice behavior of participants. The effect has been demonstrated with various tasks however the most famous one is the Asian Disease Task by Tverksy and Kahneman themselves. Additionally to the variations in task texts even the changes of outcomes from hypothetical to real ones with large and small amounts of money could not significantly influence stability of the effect (Kühberger, Schulte-Mecklenbeck, & Perner, 2002).

Given the large attention framing received over the years the factor of *comparable versus non-comparable* choices did not receive a large amount of attention. Johnson (1984, as cited in Payne et al., 1993) nevertheless revealed an interesting effect in the area of consumer research. The author presented non-comparable alternatives like a vacation trip versus the situation of buying a car. In order to overcome these non-comparable situations consumers tended to generate *higher level abstractions* (e.g., necessity of a good) to be able to make the options comparable again. This allowed them to break choices down onto the level of comparable attributes again.

Another examples deals with the *correlation* among *attributes* of a choice set. A feature of a choice is a *positive correlation* between attributes when alternatives are similar. On the other hand, a very dissimilar choice set should result in a *negative attribute correlation*. There seems to be a strong impact of correlation structure of a task and decision behavior. Bettman, Johnson, Luce and Payne (1992, as cited in Payne et al., 1993) find that with a negatively correlated set, participants turn to a more alternative-based, increased processing and less selective strategy. This point was more recently addressed in Fasolo, Misuraca, and McClelland (2003) who wanted to demonstrate the usage of decision aids in multi-attribute consumer decisions. Lists of cameras were modified into either positively or negatively correlated multi-attribute sets. With a MouseLab like tool (programmed in Javascript to be able to run experiments on the Internet), the authors found that with positive correlations, unaided choice processes were more attribute-based and choices easy. Whereas with negative correlations unaided choice processes were more alternative-based and choices difficult. In terms of compensatory versus non-compensatory aiding factors, non-compensatory decision situations performed better with positive correlations whereas compensatory sites perform better with negative correlations.

As a next step, away from the very basic assumptions on heuristics, task- and context-effects two factors: *accuracy* and *effort* of a decision are central in the ADM theory.

4.3.2 The accuracy-effort framework

Payne et al. (1993) lay out a framework based on the *accuracy* a decision maker wants to achieve and the *effort* he is willing to invest for this goal. Five assumptions are made by the authors:

1. People have a *repertoire* of strategies or heuristics in order to solve a decision problem. The acquisition of these heuristics is either possible through training or through "natural acquisition".
2. A strategy is always connected to certain *advantages* and *costs*. These two shape the selection process.
3. The structure of the *environment* promotes or demotes the application of strategies. This assumption contains the relative attractiveness of the same strategy in different environments. One strategy with excellent attractiveness ratings in environment A might perform poorly in environment B.
4. An optimization criterion is set as the fourth assumption stating that the individual chooses the *best anticipated strategy* for a given environment.
5. Through the application of a *top-down process* the decision maker is supposed to select a strategy by considering information about the task. Advantages and disadvantages are weighted an the best remaining strategy chosen. Payne et al. (1993) mention a second possible process

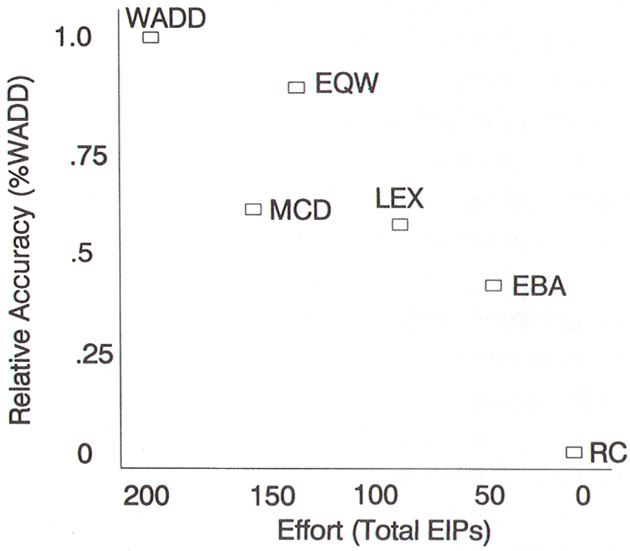


Fig. 10. Effort and accuracy levels of decision making heuristics. Source: Payne, Bettman & Johnson, 1993, p. 93

working bottom-up that learns from data without including prior knowledge. An "on the fly" selection or generation of strategies would be possible using this approach.

Summing up these assumptions, the following picture emerges: A decision maker should realize, either from the environment (the task) or from the data, what strategy fits best to a situation. This realization contains a weighing of accuracy and effort between the available strategies.

Payne et al. (1993) used two approaches to verify these assumptions. The first approach included a *theoretical analysis of effort* necessary for different strategies with subsequent simulations of these strategies. The second approach used MouseLab for gathering behavioral data.

Simulation

In a number of simulation studies, Payne et al. (1993) tried to differentiate heuristics concerning their effort and relative accuracy (i.e. how did the heuristics perform in comparison to a normative rule – the WADD model.

The relative accuracy measures of the strategies were calculated using the following formula:

$$RelativeAccuracy = \frac{EV_{heuristicsrulechoice} - EV_{randomrulechoice}}{EV_{expectedvaluechoice} - EV_{randomrulechoice}} \quad (1)$$

The second dimension – effort – was estimated using the concept of *elementary information processes (EIPs)*. The idea behind EIPs goes back to Newell and Simon (1972, as cited in Payne et al., 1993, p. 77) and states that every operation in the decision making process can be decomposed into elementary mental operations like READ, ADD, CHOOSE ... . With these, the description of a wide range of processes is possible.

Figure 10 depicts the above introduced decision heuristics in terms of their relative accuracy (calculated with Formula 1 on the y-axis and their effort (how many EIPs to consider) on the x-axis) for a hypothetical example. The endpoints of each axis are marked by WADD (highest

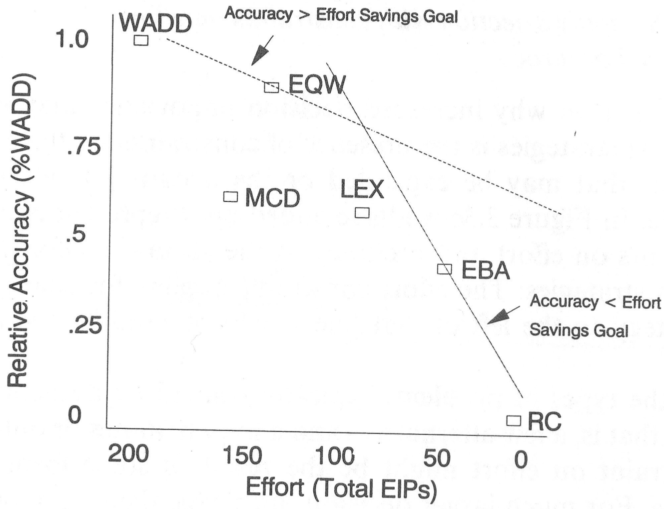


Fig. 11. Selections using effort and accuracy as criteria. Source: Payne, Bettman & Johnson, 1993, p. 95

accuracy, highest effort) and RC (lowest accuracy and lowest effort). The heuristics accumulate in the middle with EBA still under 0.5 in accuracy but with very low effort, LEX and MCD marginally above 0.5 accuracy and double the effort of EBA. EQW reaches the highest accuracy value and still performs better on effort than MCD but not LEX and EBA.

A hypothetical selection process of a strategy is depicted in Figure 11. The two dotted lines represent two selection criteria[2] – in the upper half the criterium with more weight on accuracy than on effort, in the lower half the criterium with less weight on accuracy and more on effort. The most outward intersection between a selection criterion and a strategy marks the appropriate strategy for this line. This is *EQW for accuracy > effort* and *EBA for accuracy < effort*.

Payne et al. (1993) hypothesize that different (hypothetical) environments (tasks) might have varying influences on strategy selection. Figure 12 suggests two environments (A and B) which could e.g., differ in number of important attributes for a decision. In the lower preference function EQW would be selected with high accuracy ratings and still acceptable effort. Whereas in the upper preference function (environment B), LEX rule clearly dominates, being nearly as accurate as WADD while only using half the effort.

To evaluate these predictions, Payne et al. (1993) ran a simulation study using six heuristics (EQW, SAT, MCD, LEX, LEXSEMI, EBA) versus WADD as the normative counterpart. Additionally, a combination of EBA and WADD as well as EBA and MCD was inspected. In the simulation the *task* variables, the number of attributes (two, five and eight), the number of alternatives (two, five and eight) as well as the time pressure were varied. The strategies were tested on 200 randomly generated decision problems with a count of choice and EIPs after each single trial (and the summing up of these figures afterwards). A low and high probability dispersion marked two environments the strategies had to compete in.

For the analysis, the relative accuracy of a strategy compared to the normative WADD (see Formula 1) was calculated. Regarding performance of strategies in different contexts a large variation was found between the decision environments (see Figure 13). The largest "gain" through minimizing effort (only 40% of WADD) while still performing at 90% accuracy was reached with

[2] a linear function has been chosen by the authors in this example for simplicity

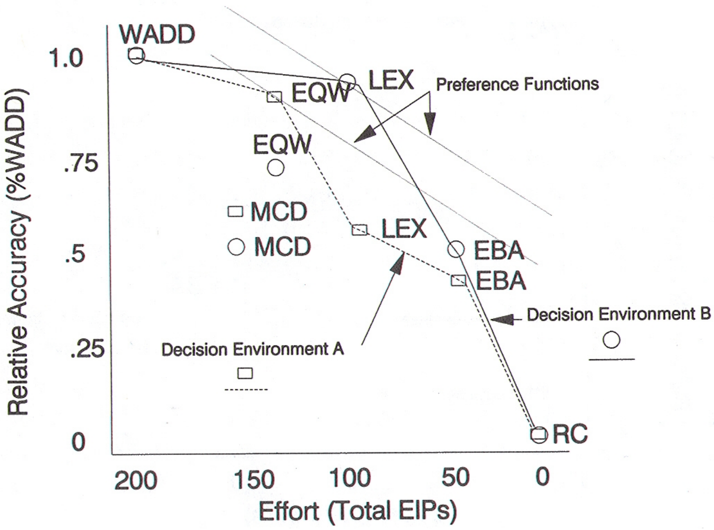


Fig. 12. Selection criteria in two different environments. Source: Payne, Bettman & Johnson, 1993, p. 98

LEX. A close follow up was EQW with 89% accuracy and 50% effort. LEX reached this value in the high dispersion whereas EQW was the best performing strategy in the low dispersion environment. From the two combined strategies, EBA+WADD performed well. With a relatively large number of EIP, EBA+MCD showed no improvement in accuracy through the combinations. It seems important which strategy combination are executed to gain best results. In the first combination – WADD+EBA – WADD rule's high accuracy can be preserved and EBA's good performance on relative effort help in a resulting score that is only outperformed by EQW, LEX and LEXSEMI. The second combination – EBA+MCD – is only in the midfield concerning accuracy but results in a good performance on the effort dimension. What can be learned from the combination of strategies is the *tradeoff* on one dimension that is always resulting through such a combination. The application of only one strategy always did better and is psychologically more plausible than a combination of two.

On the alternative X attribute dimension, the simulations showed that an increase in *alternatives* did not influence the heuristics accuracy a lot, whereas a larger number of *attributes* resulted in a decrease in accuracy. The picture is even clearer when the focus is set on EIPs. Generally with an increase in number of alternatives or attributes, an increase in the number of necessary operations could be observed.

Behavioral Data

The results from the introduced simulations were compared to studies using real data. Two of these studies from Payne et al. (1993) will be briefly discussed now. An examination of sensitivity of the decision maker to goals (a focus on attributes or a focus on alternatives), a variation in time pressure and a context variable (probability dispersion) were varied. In both studies the MouseLab system (see Chapter 3.2, p. 13) was used as the preferred method.

The drawing attention of the participants to accuracy dramatically changed their tracing behavior. More processing (large number of looked at information and more time spent on information) was recorded when maximizing accuracy was demanded as well as more alternative-wise searches. Translated to the accuracy-effort hypothesis, the emphasizing of accuracy results in a more *normative* behavior from participants.

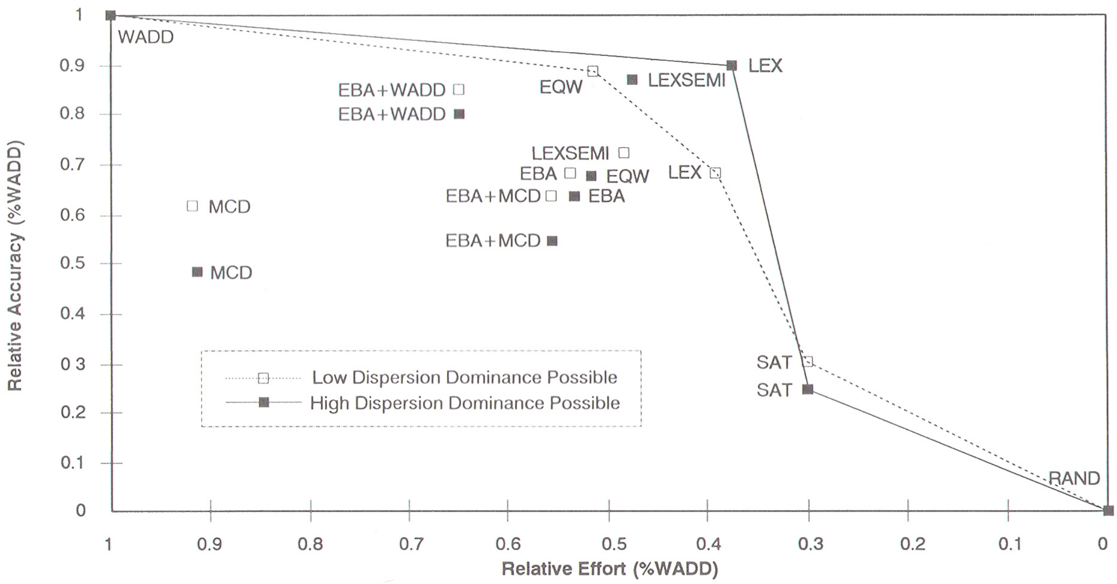


Fig. 13. Simulation of effort-accuracy tradeoff. Source: Payne, Bettman & Johnson, 1993, p. 132

When probability dispersion and time pressure were varied in the simulations, a more attribute-wise, probabilities- and “most important attribute”-focused processing in the high dispersion environments were found. In these real experiments, dispersion was crossed with time pressure resulting in adaptation (but not in all observed cases). Less information was processed in an attribute-wise way when dispersion was high. Concerning time pressure, a decrease in information processing was observed while the time spent on an item was minimized. Information search was more selective (more important items were preferred) and clearly attribute-based. This picture was accentuated when time pressure was higher (*15s* versus *25s* condition).

The ADM approach introduces a program that is based on theoretical considerations (the simulations introduced above) as well as a multitude of behavioral data (see, e.g., Chapter 3.2, p. 13 and Chapter 3.3, p. 14). In this approach, two concepts – the accuracy of a decision and the effort to achieve a decision – are of special interest. The combination of these two measures with the MouseLab system enabled Payne et al. (1993) to reduce decision situations to a level of complexity in which relatively broad conclusions and predictions of behavior are possible. One additional measure which is often used in the ADM approach – the Payne Index – is introduced next.

4.3.3 Indices for process tracing

There are several ways of representing information gathering processes in data. The following type of analysis was used by Payne et al. (1993) as well as many other researchers (e.g., Fasolo et al., 2003; Newell, Weston, & Shanks, 2003; Chu & Spires, 2003). Some critique will be addressed which was raised because of a overestimation of two types of transition (Type II and III, see Table 2) while ignoring others (type I and IV).

The Payne Index

One way of describing information acquisition in an information board task is to record whether the participant's search tended to go within- or across-alternatives. An index to compute such a measure was introduced by (Payne, 1976). Assuming that the total number of within-alternative transitions is $N_{alternative}$ and the total number of across-attribute moves is $N_{attribute}$. Payne's index (PI) is then the ratio:

$$PI = \frac{N_{alternative} - N_{attribute}}{N_{alternative} + N_{attribute}} \tag{2}$$

This index is equal to 1.00 if $N_{attribute} = 0$, and equals -1.00 if $N_{alternative} = 0$. If there are equal numbers of both types of transitions, the index equals 0.0. Using this calculation, a positive index value indicates that the search is more *alternative based*, whereas a negative index value indicates a more *attribute based* search.

For the predictions of a two phased decision process, Stokmans (1992) formulates a need for an addition to the PI. The author points out that there are four possible transition when information search is inspected (see Table 2 and Figure 14)

Table 2. Types of transitions in an information search task

Alternative	Attribute	
	Same	Different
Same	Type I	Type II
Different	Type III	Type IV

A *Type II* transition corresponds to an alternative-wise (in Payne et al., 1993's approach) search pattern, whereas a *Type III* transition is attribute based. The PI is calculated using these two types of transitions (see Formula 2). However Stokmans (1992) argues that *Type I* (i.e., no transition – the same information item is inspected again) and *Type IV* transitions (i.e., a diagonal transition with switching of alternative as well as attribute) are ignored in this approach. The author introduces a modification of PI developed by Van Raaij (1977, as cited in Stokmans, 1992).

The Van Raaij Index

The same data as in the PI are used for calculation of this index. However an additional feature is the splitting of the sample and the subsequent comparison of transitions between the first and the second half of the search process.

$$VR = \frac{N(typej)_1 - N(typej)_2}{(M-1)} \tag{3}$$

in which:
N = number of observations for a type of transition
j = type of transition (type II or III)
subscripts 1 and 2 represent the two phases of the decision process
M = total number of information items searched for

Stokmans (1992) compared the performance of the two indices (PI and VR) on the same data set. An improved sensitivity for detecting differences in strategies was observed for VR. PI identified only 6 out of 18 possible transitions in the test data, whereas for VR this number was 14 out of 18. The author concluded that the comparison of both indices is important to get a clear picture of a given data set.

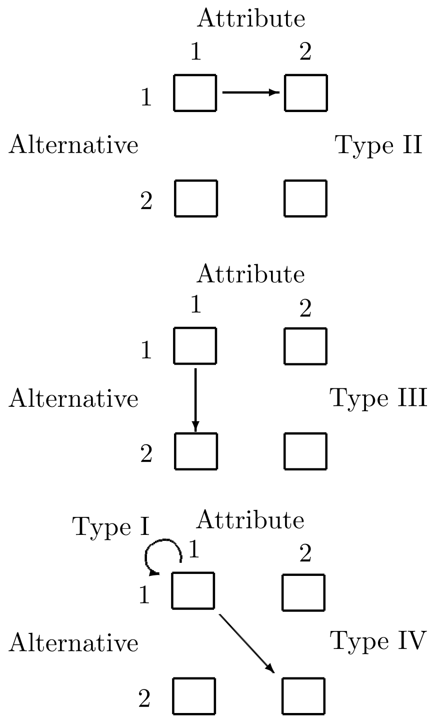


Fig. 14. Types of information-selection

4.4 Summary of approaches

Different theoretical perspectives on decision making, all with a focus on process tracing have been examined in this chapter. In many studies, not only the process is central but additional hints at either domain/task effects or effects on the person are incorporated. Table 3 summarizes the theories and their predictions of how the decision making process is happening.

Looking at the main problem of how to *describe* a process accurately, the three theories have similarities but also limitations in their approaches. H. Montgomery and Svenson (1989a)'s SDS theory searches for a dominating alternative which is then compared to all other alternatives. If the requirement of dominance can be assured for the two left over alternatives choice can easily be performed between these two. Svenson (1996)'s DiffCon theory suggests that the differentiation process between alternatives is performed before as well as after a decision. Before the decision, a *sufficiently* superior alternative is differentiated. After the decision, a *consolidation* between the chosen alternative compared to the not-chosen alternative is predicted. This consolidation can be observed when preference ratings disperse. Finally, Payne et al. (1993) suggest that the editing of information could happen anytime during the decision process (which is a kind of dominance structuring, too). Central to the ADM approach is the introduction of an *effort-accuracy* tradeoff during the decision process. The amount of alternatives and attributes or time pressure and dispersion of information in a decision situation change parameters like e.g., considered information or direction of search.

The central concept of SDS and DiffCon, *dominance*, and the *accuracy-effort* concept of the ADM are closely connected. In a decision about dominance, the final decision is always between two alternatives (a choice that Gigerenzer et al., 1999 would call "frugal"). This can also be found in the ADM approach where a reduction of alternatives is demanded for an accurate decision.

Table 3. Key concepts of used approaches

Theory	Authors	Decision Process
SDS	H. Montgomery (1989)	search for a dominating alternative; test if alternative holds against other alternatives; in most of the cases: reevaluate the alternatives
DiffCon	Svenson (1992)	eliminate alternatives to find dominating one; if one alternative is superior – choose it, if not – go back; after the decision – consolidate choice
ADM	Payne et al. (1993)	decision strategies are sequences of mental operations; in an accuracy effort-framework, the decision maker decides "on the fly" which strategy seems promising

Conceptual ideas of all three theoretical approaches will be used in the second study (see Chapter 8, p. 77) and discussed in terms of applicability to the current data.

5

Decisions modes and decision domains

The aim of this section is a classification of decision modes and decision domains. This aim seems to be an easy one. However, after closer inspection it turns out that it does not only involve decision making research. Entire research areas touch the subject of classification – examples are: very basic psychological research like problem solving in different domains, statistical methods, informatics, library- or information-sciences.

Frederiksen (1972) claimed the need for a *taxonomy of situations* in order to be able to predict behavior of individuals. As an example, the author describes Cattell's (1946, as cited in Frederiksen, 1972) use of a dictionary to find words which were descriptive for personalities. Unfortunately, dictionary entries for decision situations are not readily available. Looking for a more general approach, Frederiksen extracted three factors important for a classification: the *situation*, the *person* and the *elicited behavior*. In order to be able to draw valid conclusions with this approach, behavioral data from a large number of persons, in many situations, are required.

In this chapter, a look at definitions of classifications, taxonomies and typologies is presented first. Each of these termini is used in literature dealing with, e.g., domain differences, content effects or decision modes and should therefore be evaluated more closely. Following this part, the literature on decision modes will be inspected and the identified concepts classified within a cluster analysis. After this step, the remaining selection deals with domain differences and highlights results of a survey which tried to identify decision domains. This chapter's overall aim is to serve as a basis for the second study (Chapter 8, p. 77).

5.1 What are classifications, taxonomies and typologies?

Bailey (1994) defines *classification* "... as the ordering of entities into groups or classes on the basis of their similarity" (Bailey, 1994, p. 1). Classifications are either *uni-dimensional* being based on only one dimension or *multi-dimensional* being based on a number of correlated/related dimensions. From this general form, (1) *typologies* which are generally multidimensional and conceptual (i.e., provide concepts/names for each available cell) can be separated. When the number of dimensions and categories grows, it is hard to define every single cell. Therefore *partial typologies* are often used. A second distinction (which can be found in the decision making literature, too) is (2) *taxonomy*. Taxonomies describe either the process (of generation) or the end result of a classification. They differ from typologies in providing a classification of *empirical entities* instead of conceptual ones. However taxonomies and typologies are often used interchangeably.

Classifications bear the following *advantages*: they (1) *describe* all necessary dimensions in a clearly arranged form; they (2) *reduce complexity* and therefore (3) make it easy to identify *similarities* and *differences* between cases. Furthermore the (4) *comparison of types* as well as the (5) *study of relationships* is possible. However certain *disadvantages* are discussed by Bailey (1994) like the (1) *unmanageability* of large classifications as well as the oversimplification of small ones. The (2) *identification of cases and variables* is a very basic problem applicable to nearly every approach in the social sciences – the better the variables the better the quality of the outcome. This is also true for classifications. Another point of critique mentioned by Bailey is the (3) *descriptiveness* of the approach and therefore the *pre-theoretical* level a classification always has. This last point is relativized by the author by pointing out that the explanative value of a theory always depends on the quality of the classification system underlying it – without a good system only weak assumptions can be drawn.

Bailey (1994)'s three-level model seems especially applicable to the current situation of research on domain differences in decision making. The author differentiates the following levels of measurement: the *conceptual*, the *empirical* and the *indicator* level. The *conceptual level* contains typologies of a conceptual type with no *empirical counterparts*. In contrast to this, the empirical level is always derived from empirical measurement. Both types can occur in a "pure" form as completely conceptual (theoretical assumptions about types) or as completely empirical (measurement without theoretical basis). The third, the indicator level, emerges when a transition between level one (conceptual) and level two (empirical) or vice versa is done. It describes the process of forming a concept and then looking for empirical cases, or clustering empirical results first and formulating conceptual labels afterwards.

Following this very general distinction from the decision making point of view, several terms are used to describe decisions from different classes. The usage of these is interchangeable but especially two terms need differentiation: the *domain of a decision* and the *decisions mode*. The *decision domain* can be described best as a typology of decision situations, e.g., financial versus medical versus juridical. The *decision mode* represents miscellaneous states a decision can be performed in like e.g., normative, emotional or rule based.

To be able to build a classification, of decision modes and decision domains two sources were consulted. Maule and Pearman (1994) for the literature between 1974 to 1994 as well as Gilgen and Schwizer (2003) for a more recent collection. In both sources, extensive literature reviews were performed. Maule and Pearman (1994) additionally tried to build a taxonomy whereas Gilgen and Schwizer (2003) concentrated on a broader *search* result. I will first review the literature chronologically and will then concentrate on the effect of *decision content* (which is used only in very few examples).

5.2 Decision modes, a large table and a clustering study

The old days

A literature review of decision making literature dealing with different decision modes of the years 1974 to 1993 was performed by Maule and Pearman (1994). Despite the intended usage of this report in a military context, an overview of classifications in the areas of individual, group and organizational decision making is given (for the purposes of this thesis, only the individual decision making part is taken into account). The literature described by the authors will be incorporated in more detail in Table 5, p. 46.

Six classes of decision making modes are extracted:

1. automatic/intuitive with little or no prior knowledge
2. automatic/intuitive based on prior knowledge
3. analytic

4. rule based with rules applied in many situations
5. rule based with rules applied only in one situation
6. action or learning based

- Class 1 and 2 concern *automatic* and *intuitive* decisions. These are described as being decisions of low importance, with little control, many attributes, low involvement and little differentiation between attributes. Arroba (1977) summarized this behavior into a *no thought* category pinpointing at the automatized part of this decision mode. An example for a decision situation within class 1 would be the stopping at a red light. For the second class, managerial decision making which is often highly rated in the public can serve as an example. Maule and Pearman (1994) saw a further need for differentiation of the *no thought* category. *Automatic processing* in lay people should receive another class than that of experts. The automatization in expert decision making (which should be larger with an increasing of expertise) should not reflect a general tendency of intuitive decision making being solely that of an expert. In this line of research, the fact of expertise therefore is not a prerequirement for automatic processes.
- Class 3 – the analytic – defines the other end of the continuum with *high cognitive control*, awareness, slow information processing, high involvement, good differentiation between alternatives and lower confidence in the final result (compared to the automatic class). This class represents decision making research's main field of interest despite a multitude of studies show that, e.g., heuristics are used instead of purely analytic strategies. As an example for this class, which is often used in decision aids, is the application of an Multi Attribute Utility Theory (MAUT) approach. Using this method for each attribute of a decision problem weights and utilities are multiplied and compared among each other. An often used example for a MAUT approach is the decision about which computer or which mobile to buy. Important about these examples is the independence of attributes, which might not always be given (a small price can be a tradeoff for less disc size or processor speed, in the computer approach).
- Class 4 and 5 decisions have in common that *rules* are involved. In class 4, these rules are *general* – for many situations – "a bird in hand is worth two in the bush". In class 5 these rules are *content specific* – applicable only in one situation – e.g., wear a warm sweater when it is cold outside. In these situations, the analysis of the decision situation can be reduced to the identification of a trigger. When a trigger is recognized the corresponding rule is executed.
- Class 6 was introduced by Reid and Crompton (1993, as cited in Maule & Pearman, 1994, p. 30) especially for leisure purchase decisions. It describes a decision where first an action is set and the outcome of this action is then taken as the basis for the next decision Unique to this approach is the pre-decisional process of "acting first and deciding later". Nevertheless, this idea seems to be bound to quite a small number of situations and only of little interest to the current research idea.

Summing up these six classes, a division into three parts seems natural. These parts are an *analytic-* versus an *automatic/intuitive-* versus a *rule based*-one. Contrasting analytic with intuitive and rule based decision modes is a practise often used when it comes to classifications of these.

Recent Years

A literature search performed by Gilgen and Schwizer (2003) covers the years between 1993 to 2003. In this search, the following databases were included: ALEPH, PsycINFO and scirus.com. Keywords used were[1]:

[1] a * indicates the usage of a "token" – any letter-digit combination can be put there

decision taxonom*, decision classification*, decision types, decision categor*, decision classes, decision domain*, classifying decision, classes of decision making, domain* of decision, taxonom* of decision making, classification of decision making, type* of decision making, categor* of decision making, domain* of decision making, typology of human decision making, individual decision making, underlying processes AND decision making, topic of decision, domain of decision, content effect AND decision making, content of decision

Central results of this search were the papers of Blais and Weber (2001), Medin, Schwartz, Blok, and Birnbaum (1999) and Weber, Blais, and Betz (2002). These papers are now examined in more depth:

The main interest of Blais and Weber (2001) was to combine various decision modes with gender differences. Because gender differences are of no interest to this study, only results concerning content effects will be explored. The authors used short stories describing different situations with two courses of action each (a risky and a save one). For each situation five reaction modes were presented and participants were asked to assign one mode to one situation each (see Table 4).

Table 4. Decision Modes used in Blais and Weber (2001)

Mode	Description	Actual option
Compromise	compromise between costs and benefits	I try to find a compromise
Authority	relying on advice of respected others	I follow my friend's advice
Status-quo	avoiding any action	I postpone the decision
Rational	personal utility standpoint	I weigh the pros/cons of the situation
Emotional	intuition, emotions, gut feelings	I go with my feelings

Asking for advice (the *authority* mode) showed the largest variations between the situations. In three situations ("break up relationship", "choose graduate school" and "perform birth control") participants favored this mode, whereas in the other two situations ("car purchase" and "cheating on term paper") it was the least favored mode. The other decision modes also varied between stories but not to the extent of authority mode. Risk behavior was examined through the actual decisions participants wanted to perform. Here clearly the "cheating on term paper" and the "perform birth control" scenario elicited the choice of the save option. In contrast to this, the "car purchase" scenario showed nearly a 50:50 distribution between save and risky option. Summing up these results: reducing the response options to just five decision modes and varying the decision situations results in a clear cut and well interpretable picture of differences in behavior when domains (tasks) differed.

Medin et al. (1999) set a much broader focus on decision modes, i.e., the meaning of a decision to the decision maker. In three experimental conditions the authors varied the meaning an object could have to the participants. This was done in terms of three factors that affected exchangeability: *sentimental value* (a story was constructed where land, that was owned by the decision maker's family for some time should be sold; higher sentimental value was expected if the land was owned by the family longer), *monetary value* (the cover story introduced a person who owned a dog for some years and is asked to sell it, in one condition, to a hospice, in the other condition to a stranger) and *source-independent value* (in this scenario the background of a book was varied, in one condition, once owned by a bookstore, in the other condition by a famous person). Each of these scenarios were constructed and run in separated experimental conditions. The results demonstrated that meaning influenced judgments and attitudes equally. On the dimension of *sentimental value*, time of ownership and intended use of a good had a large influence on the stated minimum value for exchange of the good. On the dimension of *monetary value*, a strong influence of intended use of the “to be sold” good (the dog) was found. Finally the *source-independent value* condition produced the high resell rate if, e.g., a famous person once owned an item of interest compared to an anonymous bookstore as an owner.

For explaining the differences between these tasks, Thaler (1985)’s mental accounting idea was used. It states that accounts exist to be able to evaluate expenditures and returns over time. Accounting also enables the decision maker to spend resources to new problems because old ones already got an account and need no further processing. As a third factor accounting enables learning through considering unsuccessful decision attempts and changing behavior afterwards. The (not very surprising) conclusion of Medin et al. (1999) is that decision making is “infused” with meaning - the “how” and “to what extent” is not answered in this paper. Some criticisms on the method must however be mentioned: it does not become clear from the experimental setting what exactly influenced the changes in judgment. There was no check whether the two compared conditions are still on the same dimension, i.e., is the sentimental value of a children’s hospital comparable to a shopping mall? Furthermore there seems to be an overlap at least between the sentimental and monetary value condition. Clearly Rettinger and Hastie (2001)’s method of keeping the expected value of all the options constant seems more promising (see Chapter 5.2.1 for details).

Weber et al. (2002) developed a scale for assessing risk attitudes and tested it in five domains: financial, health/safety, recreational, ethical and social decisions. New in their approach is the differentiation between *risk attitude* (in the classical sense of behavioral decision theory, i.e., the “curvature” of the utility function) and *perceived risk attitude*. The later takes into account personal and situational factors, i.e., a person can be risk seeking in one situation and risk averse in another – but still have the same risk attitude. Domain specificity of risk taking is accredited to differences in the perception of choice alternatives (in different domains). For the above five domains, an easy to use instrument was generated which gives reliable information of a person’s attitude concerning risk from different points of view (domains). This instrument lents itself to group people according to their perception of risk and analyze data according to these groups. An important difference between Weber et al. (2002)’s study and the current thesis is that risk attitude (in Weber et al., 2002) and actual risk taking behavior (the topic of this thesis) result in different behavioral reactions. Risk attitude comprises more or less intention to do something, whereas risk taking behavior has to include the actual action.

The experimental setup participants are confronted with is crucial in the first two papers above (Blais & Weber, 2001; Medin et al., 1999) and specific for research on domain differences is. Most often, a (decision) situation is described and participants are asked to rate or judge how they would behave in such a situation. Blais and Weber (2001) gave participants one of five decision modes as options. Participants then had to choose one of the five modes for a number of decision situations. Medin et al. (1999) presented different stories, too. The authors examined decision between them (as already mentioned above, Weber et al., 2002’s study is different in assessing *risk attitudes* in contrast to *risk behavior*).

The central concern about these approaches is the value induction without control what the actual meaning of a certain value is to the participant. Using various scenarios and identifying differences between them leaves a problem in the interpretation of the resulting data. Rettinger and Hastie (2001) provide the only study, the author is aware of, that controls for equality of values between tasks through changing content but keeping probability and outcome constant. This idea seems to be extremely important if domains want to be compared.

5.2.1 Decision modes and taxonomy generation

To be able to draw hypotheses about decisions in different domains in current (and not so current) literature – a two step process is suggested. Step one lists the relevant literature dealing with decision domains and decision taxonomies. Here Maule and Pearman (1994) and Gilgen and Schwizer (2003) are taken as starting points and are extended with additional sources.

Table 5: Decision Modes

Source	Mode	Action
Arroba (1977)	logical	what is objectively the best alternative
	no-thought	frequently encountered decision which has low importance
	intuitive	personal feeling of rightness
	emotional	choice is based on subjective preferences and feelings
	hesitant	postponing a decision
	compliant	choices according to other people's perceived expectations
Blais and Weber (2001)	compromise mode	cost and benefits are considered → compromises and trade-offs between the two are used for the decision
	authority mode	advice of respected other is considered
	avoidance mode	avoiding of any action and remaining at the status-quo
	rational mode	generating pros and cons → selecting overall best option
	emotional	intuition, emotions and gut feelings are key components
Goldstein and Weber (1995)	nondeliberative decision making	repeated and routinized decisions
	associative deliberation	
	rule-based deliberation	following a plan for making the decision
	schema-based deliberation	construct model → generate choice alternatives → when they fit, make decision
	social relationship	constructing stories/schemas
	inanimate objects	weighing advantages and disadvantages
Klein (1993)	recognition primed	recognize and classify a situation → react in "typical" way; if there is time: simulate outcome (expert decision making)
	analytic	used if data are abstract and action has to be justified
Loewenstein, Weber, Hsee, and Welch (2001)	intuitive	emotion dictates the decision without the influence of cognition
Maule and Pearman (1994)	automatic / intuitive with little or no prior knowledge	low in awareness and cognitive control
	automatic / intuitive based on prior knowledge	automatic decisions based on expert knowledge
	analytic	high cognitive control, high involvement and cognitive analysis
	rule based with rules applied in many situations	application of general rules – always choose the cheapest
	rule based with rules applied in only one situation	application of context specific rules – choose the alternative most similar to situation x
	action or learning based	after an action is set the outcome defines the ongoing decision
Medin et al. (1999)	decisions about sentimental value	the higher the sentimental value → the less willingness to sell something
	decisions with attributions to someone's goal	the more favorable a good is → the less willingness to trade it

Payne et al. (1993)	analytic strategies	through the use of heuristics many alternatives are reduced to few (two) and a decision is performed
Tada and Weber (1998)	school/professional decisions	which college shall I attend
	relationship decisions	friendship related
	group involving decisions (non deliberative)	whether to cry
	group involving decisions (value laden)	religious preferences, drug use
	monetary and consumer decisions	whether to buy certain things or not
Wagenaar, Keren, and Lichtenstein (1988)	deep structure	choice among bet options
	surface structure	content of decision or position in decision situation
Weber and Hsee (1998)	cost benefit / calculation based	weigh and combine the likelihood and desirability of outcomes
	recognition-rule-based	recognize a situation → apply the stored rule
	recognition-role-based	e.g., parental decisions for their children (strengthening social identity)
	recognition-case-based	remember a situation in the past → take action (expert decision making)
	affect-based	affective reaction to different choice alternatives
	nondeliberative	routinized decision, if there is a red light → stop
	stereotype-based	retrieval of stored judgment (e.g., opinion on a group)
	principle-based	if offered some illegal drugs → just say no (affect-based decisions may lead to bad outcomes
	schema-based	construction of alternative stories → evaluating of what might happen
Weber et al. (2002)	health / safety decisions	seatbelt usage
	recreational decisions	skydiving versus bowling
	social decisions	confronting co-workers
	ethics decisions	cheating on exams
	investment decisions	investment in a stock
	gambling decisions	betting on horses
Yates and Lee (1996)	analytic	identify options and chances of outcomes → assess values
	rule based	build up collection of rules → match current situation to one in rule store
	automatic	automatic matching of situation with rules

The results of this step can be found in Table 5. Step two reduces these different domains to their least common denominator (see details in Chapter 5.2.2). As Table 5 illustrates, there is a large number of concepts of different decision making domains. To be able to categorize theses concepts a card sorting study, followed by a cluster analysis was run.

5.2.2 Clustering study on decision modes

The aim of this clustering study was to build a classification of decision modes from the identified literature.

Participants.

Experts in decision making and/or Psychology, as well as lay persons (N = 23) were recruited from the University of Fribourg. Of this group, 10 were female with a mean age of 25.00 (SD= 2.91 years) and 13 male with a mean age of 29.08 (SD= 5.82 years). Participants received the descriptions of decision making domains found in the literature (N=56, a list of these can be found in the Appendix, p. 115) and were asked to group them into categories.

Method

The task consisted of three steps. Step 1: make groups from the source list; step 2: combine the groups into higher level ones and finally step 3: give a name to the higher level groups (detailed instructions for every step can be found in the Appendix, p. 116)

Apparatus.

A Pentium 4 desktop PC with Windows XP and a 17" screen was provided. USort (Dong, Martin, & Waldo, 1999), a computer program written to run card sorting experiments, was used to gather the data. Although this program was intentionally written for the evaluation of Web-sites, USort is a convenient tool for finding underlying structures in unsorted data.

On the initial screen, participants saw a list with the 56 items (see Appendix, p. 115). Through drag and drop the sorting into groups was possible. An already sorted item could be regrouped again during this first stage. Before participants continued with stage two, they were reminded to check whether their sorting results were satisfying. In stage two, the creating of higher level groups was possible again, through drag and drop, but now of groups. Finally in stage three participants were asked to assign names to every higher level group from step two.

Results

Data were analyzed with EZCalc (Dong et al., 1999) a program that generates *tree diagrams* that indicate the strength of the perceived relationships between pairs of card (through a cluster analysis) and displays these relationships graphically.

Five clusters resulted from the analysis which were named after the most frequent concept in each of them.

1. avoid
2. rule-based, analytic
3. automatic
4. emotional, intuitive
5. social

Cluster 1 (avoid) and 3 (automatic) (see Appendix Figure 34, p. 120) are of no interest to the following experiments because they both represent decision situations without higher cognitive processes. In the automatic, nondeliberative cluster, all decisions that require little or no cognitive load are gathered. Again information search in this cluster seems to be quite restricted. As a single outlier health decision making was grouped into cluster 3. It is one of two decision modes not sorted according to theoretical predictions (the second one is "health" in cluster 3, see Appendix Figure 34, p. 120).

Cluster 2 (rule-, analytic) which is the largest resulting cluster contains three pathes of interest. Path 1 ends in "decisions about the surface structure" – a concept by Wagenaar et al. (1988). This concept was not associated with another card in the sorting process. Therefore no cluster is resulting. Path 2 ends in a rule-based area whereas path 3 in an analytic one. It is suggested to divide the very large cluster two into two parts a rule-based and an analytic one.

Therefore the list of clusters is reduced to:

1. rule-based
2. analytic
3. emotional, intuitive
4. social

A central conclusion from the reviewed literature is that there are two very different concepts which are mixed in the discussion about decisions in different areas. The first (already described) group is the *decision mode*, i.e., *how* a decision is performed given various situations. The second (described below) group is the *decision domain*, i.e., the content a decision is about.

5.3 Decision domains

Goldstein and Weber (1995)'s perspective on content differences does not start in decision making but takes a general look at Psychology first. Numerous examples are given in which content is incorporated and declared as important, e.g., *learning* (Ebbginhaus, 1964 as cited in Goldstein & Weber, 1995), *deductive reasoning* (Wason & Shapiro, 1971 as cited in Goldstein & Weber, 1995) or *problem solving and expertise* (Anderson, 1987 as cited in Goldstein & Weber, 1995). The authors see a reversed trend in decision making research where only indirect usage of content is documented. Examples for such areas are *framing* (Wagenaar et al., 1988 seems to be a contradicting example because explicit content differences are investigated using framing tasks) and *missing information* (addressing the question of what inferences are drawn when information is degraded or missing). These two and other decision making areas have in common that content is recognized as important but seldom taken as an independent measure.

In two experiments, judgments of stimuli from different domains and importance of a decision were manipulated. Goldstein and Weber (1995) argue that a condition which induces schemata (stories, pictures) should elicit different judgment modes in the subsequent decision (as demonstrated in a condition where a spouse is evaluated). Contrasting this with a condition which is less vulnerable to such manipulation should show no or small differences (as shown with the evaluation of a CD player). The authors hypothesized that choosing a spouse would include selecting more narrative strategies while a numerical (normative) strategy would be more likely for buying the CD player. The interpretation of the results stated that different domains actually elicited different or no schemata and therefore used variable strategies. The idea was strengthened in a second experiment manipulating the importance of a decision in varying domains. This treatment had a smaller effect in comparison to the domain manipulation.

Rettinger and Hastie (2001) ran an experiment in which tasks from four different domains (legal, academic, financial and gambling) were compared. The basic structure as well as values and probabilities used between tasks were the same. The decision to be made was either to accept a sure loss or play a two-stage gamble. The three main measures recorded were (1) the chosen option, (2) the confidence in the decision and (3) an open question on strategies participants used, to solve the tasks. Analysis of the data showed that differences in chosen alternatives could be found. The gamble task (where the safe option dominated with 80%) differed significantly from the other three tasks (where the safe option was reduced to 35% in the financial, 46% in the academic and 52% in the legal task respectively).

This paper is the only one directly comparing the same formal structure with different cover stories/domains. Due to the closeness of that method to the one used in this thesis, the predictions are taken as central to the current work.

5.3.1 Survey on decision domains

After identifying decision modes by means of a literature search and a cluster analysis, the classification of decision domains is the next step.

Method

A one question survey was constructed using an online tool called Easy Survey Package (ESP) (see Appendix Chapter 33, p. 118 for a screen shot). ESP offers an easy to use possibility for conducting surveys and retrieving data in a spreadsheet readable format. Participants were asked to write down as many decision domains as came to their minds.

Participants.

An e-mail was sent to the mailing list http://www.sjdm.org/mailman/ of the Society for Judgment and Decision Making (SJDM). This list is frequented by decision making experts from all over the world. In order to keep the whole process of data collection as simple as possible no demographic data were recorded. Thirty-six responses were collected (from 36 unique IP addresses, which were checked for double entry). Among these responses were three that were no analyzable, therefore 33 replies were used for further analysis.

Apparatus.

The survey was setup on a Pentium 3 with 256 Mb RAM and Linux Mandrake as operating system. The PHP based ESP was used for creating and hosting the survey. Data were converted into a CSV-file (comma separated values) and analyzed with SPSS 12.5 for Windows.

Results

From the 33 analyzed replies 359 unique words (domains) were extracted and a frequency list with percentages was generated (see Appendix, p. 119). A large amount (24.8 %) of domains were only named once. A comparably large amount (27.0 %) was named two to seven times. The highest frequency of mentioning was received by the domains (listed in Table 6):

Table 6. Frequencies of domains

Frequency	Domain
19	medical
14	law
13	business
12	politics
11	personal
10	health
	finance
9	career
	consumer
8	social
	management

Because a large number of the listed domains overlap and task generation demands clear differences between domains I do suggest to merge the following domains into one: medical and health →

medical; business, finance, consumer and management → *business*; personal, career and social → *personal*. This fusion is done arbitrarily based on the consideration which domains belong together. The two domains remaining unique in their category are *law* and *politics*. This reduction results in a reduced list of domains – medical, business, personal, law and politics – for further usage.

5.4 Summary of modes and domains

In this chapter at first three termini were clarified. These were differences between a classification (ordering of entities into groups), a typology (giving names to each cell of a classification) and a taxonomy (which is a classification of empirical entities). For the current purposes a *taxonomy* is the most important concept because only empirical papers are included in the clustering process.

The second distinction introduced is between *decision mode* and *decision domain*. These two concepts are often used interchangeably. However for the current research questions it seems important to separate them. For both areas, a small survey was run which showed the following results for (1) modes and (2) domains. (1) With *decision modes* a cluster analysis identified four common modes in which a decision can be performed. These are rule-based (following some kind of rule), analytical (the very common, normative way to make a decision - most often used by decision researchers on their own), emotional, intuitive (the opposite of analytical mode - following solely a "gut" feeling when a decision is done) and social (this last mode includes, e.g., the peer groups opinion on a topic). With (2) *decision domains*, a survey resulted in eight frequent domains of which three (medical, law and business) were taken into account for further considerations.

The decision to use *domains* instead of *modes* for the following experiment was based on problems with decision modes like: How to induce a mode? How to check, whether the intended induction worked? How to guarantee that two tasks are from the same mode? To overcome these problems, experiment 4 concentrates on decision domains, solely.

6

Study 1 – Information search in the laboratory and on the Web

6.1 Introduction

As[1] outlined in Chapter 1 at first methodological questions are considered. The transfers of the basic idea of AIS (a participant is asking questions to an experimenter) into a computerized version is of central interest now. The proposed tool should include a search mechanism with subsequent presentation of the relevant information to the participant. Because it was not clear whether the computerization would influence the search process two search modes – list and keyword – were compared (within the new method) first. The *list version* presented all the available information (in the form of questions) to the participant who indicated the once of interest. The *keyword version* offered an environment where the participant entered a keyword, received a list of connected questions, chose one of interest and received an answer to it. The usage of questions (and answers) as information sources is necessary because it would not be possible to trace the decision maker if answers would be given by the system straight away. Only a "click" can be registered in a computerized approach and is therefore necessary for getting the actual information of interest. An interesting alternative to this approach would be the usage of an eye-tracking device where a "real" search engine style with the entering of a keyword and a subsequent list of links with some information preliminary information could be used.

Furthermore a second factor was varied – the location where the experiment took place. In the current study two locations were used a *Web version* was run solely over the Internet whereas a *laboratory version* was done traditionally in a laboratory at the University of Fribourg.

6.1.1 Online research

Inspecting the literature on Web studies it becomes obvious that in recent years, conducting experiments and other studies on the Web has become increasingly popular. Web-based studies are marked by several characteristics: (1) the procedure, including instructions and participants' reactions, is implemented on the computer via the Web; (2) participants may come to the respective Web site and end the procedure whenever they choose to without prejudice; (3) it is easy to access large samples; (4) motivational confounding can be detected; and (5) the experimenter has access to the number of nonparticipants. (For a detailed discussion of the advantages and disadvantages of Web-based studies, see Reips, 2000, 2002).

[1] A paper based on this chapter was published as Schulte-Mecklenbeck, M., & Huber, O. (2003). With or without the experimenter: Information search in the laboratory and on the Web. *Behavior Research Methods, Instruments, & Computers, 35* (2), 227-235.

Most Web-based studies use participants' reactions on a response scale (e.g., a yes/no or an interval scale) as a dependent variable. Studies validating laboratory-based experiments on the Web generally reveal that behavior in the two conditions differs only slightly, if at all (Krantz, Ballard, & Scher, 1997; Krantz & Dalal, 2000; McGraw, Tew, & Williams, 2000; Mehta & Sivada, 1995; Musch & Klauer, 2002; Pohl, Bender, & Lachmann, 2002).

However, there are areas of psychological research in which other dependent variables other than reactions measured on response scales are relevant. One example is the search for information by the participant. In experiments involving this type of behavior, the participant is asked to work on a task (for example, to solve a problem or make a decision). In order to complete the task, the participant has to actively search for (subjectively) relevant information. The type of information that is searched for, and the sequence in which the search takes place, are the central dependent variables of the behavioral analysis of the process. To our knowledge, no experimental comparisons of search behavior in a decision task between the Web and the laboratory are available. In this study, we investigate such a comparison with the example of risky decision making tasks and ask the following: (1) Is active search for information in a risky decision-making task identical on the Web and in the laboratory? (2) Are there differences between information acquisition methods (i.e., how a person gets information) used on the Web and in the laboratory?

In the first section of this chapter, we discuss the advantages, disadvantages, and validity of Web studies. We then present the main results of recent research on risky decision making in the laboratory and on the Web. In the second and third sections, we describe our experiments and their results. Finally, we discuss these results in the light of the theoretical outline.

Advantages and Disadvantages of Web Studies

Conducting experiments on the Web has gained much interest in recent years (Batinic et al., 2002; Birnbaum, 2000; Reips & Bosnjak, 2001). As is the case with every new method, questions of its advantages and disadvantages have arisen. Reips (2000, 2002) suggests that, among other things, online research facilitates access to samples by reaching large populations from different countries and groups, enables the delivery of the experiment to the participant, produces high statistical power due to large sample size, and offers the direct assessment of motivational factors through analysis of dropout behavior in certain conditions. However, it is also noted that with multiple submissions, it is possible for participants to do an experiment several times; there is a lack of experimental control (e.g., the experimenter has no influence on location and experimental setup); self-selection may be at work; and technical variance could occur, such as different access speeds and use of various operating systems or browsers.

Validity of Web studies

As we previously mentioned, conducting research on the Web has many distinct advantages that make it attractive to the researcher. Thus, it is necessary to find out whether or not the disadvantages actually are harmful and, if so, under what circumstances. In order to investigate the influence of measurement on the dependent variable, a study is run on the Web and compared with a study run in the laboratory. If there is no difference in the results, the Web version is considered to be valid.

Krantz and Dalal (2000) divided their review of the validity of Web studies into three types of research designs: surveys, correlational studies, and experimental studies. In a survey comparing traditional mail with e-mail delivery, Mehta and Sivada (1995) found that response rates were equivalent and mean responses comparable. In a correlational design study, Buchanan and Smith (1999) put a psychological test (the Self-Monitoring Scale; Gangestad & Snyder, 1985) on the Web. The authors ran a sample and compared the results with published results of traditional tests. Comparable internal consistency of the Web sample with the pen-and-paper test and a factorial analysis with excellent goodness-of-fit measures were the results. Finally, for experimental designs, two studies of ratings of attractiveness should be mentioned. Krantz et al. (1997) used line drawings as stimuli, whereas Epstein, Klinkenberg, Wiley, and McKinley (2001) developed the method further by using ratings of attractiveness of photographs. Both studies yielded comparable results between the laboratory and the Web groups.

Laboratory- versus Web-based experiments

There are only a few studies that have compared laboratory and Web experiments using decision-making tasks. Birnbaum (1999) and Anderhub et al. (2001) found small differences between laboratory and Web conditions using gambles as tasks. These differences were due to larger variances in the Web sample. Both studies attributed these results to differences between the two sample populations. Shavit et al. (2001) used lotteries and controlled the population sample by using students in both the Web and the laboratory conditions. The participants were more risk seeking on the Web. The general choice behavior was nevertheless comparable.

6.2 Experiment 1 – The comparison

Experiment 1 was intended to compare the use of two AIS-type tasks in the laboratory and on the Web. A computerized version of the basic and list versions is provided, thereby guaranteeing comparable experimental conditions at both locations.

6.2.1 Method

Participants.

The participants in the Web condition were recruited via e-mails to four German newsgroups (de.sci.soziologie, de.alt.umfragen, de.sci.psychologie, and de.sci.misc); the main topics of these groups are psychology, sociology, and science. From December 17, 2001 through February 17, 2002, 120 users accessed the experiment's Web page. The participants had to go through a registration process in which information regarding their e-mail address, sex, and age was requested. Immediately after registration, an e-mail with a link to the entrance page, a user name, and a password (which could be used only once) was sent to each participant. Fortyfour participants (36.44%) had to be excluded due to incomplete data (see Dropouts in the Results section for details). The remaining 76 participants (32 female, 44 male) had a $M_{Age} = 28.73$ ($SD_{Age} = 4.21$).

The participants in the laboratory condition were recruited via the student e-mail list of the University of Fribourg. Every student attending the university is registered on this list, and its use is frequent. Psychology students were excluded from participation because decision making and online research are regular psychology lecture topics at the University. Twenty-nine participants (18 female, 11 male) with a $M_{Age} = 24.56$ ($SD_{Age} = 3.51$) replied to the initial message. Several appointment times were proposed by the experimenter, and the participants were asked to choose a convenient one. Every participant completed the experiment.

Twenty coupons to an online bookstore were raffled among all of the participants as an incentive for participation.

Reips (2000)'s multiple-site entry technique was used to obtain insight into the origin of the participants of the Web sample. The idea was to generate several entry pages and hyperlinks leading to the starting page. After the experimenter checked the log files of the Web server, an analysis of the participants from different locations was possible. The present experiment was announced in two different locations: newsgroups and as a Google ad (see http://adwords.google.com for a description). Participant turnout totaled 120 for the newsgroups and, interestingly, 0 for the Google ad. It is surmised that since only a small number of page views was financed, the resulting presentation period was short.

AIS method on the Web: List and keyword versions.

In order to examine the research questions proposed, the AIS method of presenting information to the participants was used. This method does not affect information acquisition behavior in terms of distribution of question types between the list and keyword versions. Furthermore, it enables the experimenter to detect in which items of information the participant is interested.

Two versions (list and keyword) were differentiated. The list version is a computerized AIS version that lends itself to being implemented on the computer and, therefore, also on the Web. As an alternative method, we introduced the keyword version (see basic AIS version), in which the participant entered an individually selected keyword and was then confronted with a list of questions containing that keyword. The two versions are described below.

List version. The participants were confronted with a list of questions (in hyperlink form) and selected one at a time by clicking. The questions were formulated according to the following categories: probability (questions concerning the probability of the occurrence of an event), new alternative (questions concerning options not presented by the experimenter), control (questions dealing with the decision maker's control over the external event or over negative consequences), worst-case plan (questions demanding information about what can be done in case a negative event occurs), situation (questions asking for general background information not linked to a specific alternative) and consequences of a certain alternative. Control and worst-case questions are instances of questions for RDOs. The participants could click on as many questions as they wished (see Table 7 for an example of categories with questions and answers)

Table 7. Decision Categories

Category	Question	Answer
Probability	What's the probability of a new contract if I buy the machine?	The probability of a new contract is 50%.
New alternative	Can I get expert statements on two machines?	No experts are available at the moment.
Control	Do I have influence over future orders of my factory?	Future orders depend on the current assignments.
Worst-case plan	Can I get insurance against financial loss if one machine breaks down?	Yes, there is insurance against financial loss, but premiums are high.
Situation	What does the factory produce?	The factory produces mechanical parts for watches, electric motors, and electronic microcomponents.
Consequences	What are the consequences if I agree to buy a new machine?	Your employees will be able to work with the latest available technology.

Keyword version. In the keyword version, the participant entered a keyword into a database (similar to using a search engine) and decided which information from the resulting list was interesting. The participants searched through the same questions used in the list version, except that now they accomplished this by entering a keyword. A list of corresponding questions in hyperlink form was presented; when a participant clicked on a question of interest, the answer was shown.

Decision-making tasks.

A warm-up task and two main tasks were used in the experiment. As the warm-up, the post office task was used (see Table 8). The two main tasks were the ticks task and the mechanics task (see Table 8). The basic structure of all three tasks was the same.

Table 8. Decision Tasks

Task	Text
Post	Imagine that you are the head of the post office in the village H. The post office has very cramped conditions and has been faced with the following problem for several years: In November and December, it has to handle many parcels. If the number of parcels to be handled is too large, conditions may become unbearable. The village administration offers you the opportunity to rent the local meeting hall.
Ticks	Imagine that you are the director of a pediatric allergy center. It is uncertain whether the center can stay in the rented location, because the owners may sell the building. If this sale take place, the center will have to move out and the director will have to find another location that meets the requirements for treating allergies. You have just now been offered a big house in the forest. The problem is that the forester's house stands in a wood, which is contaminated with a specific kind of ticks.
Mechanics	Imagine that you are the owner of a small factory that produces a kind of weatherproof credit card for ski lifts. The order books are so full that an additional high-capacity machine must be bought to meet the great demand. You can choose between two machines. They differ in their technology and in how prone they are to interruptions caused by breakdowns.

Design.

In this study, three independent variables were examined: location (laboratory vs. Web), version (list vs. keyword), and task (ticks vs. mechanics). The independent variables of location and version were varied between participants, whereas task was varied within participants. The procedure was exactly the same for both the laboratory and the Web conditions. Both groups were instructed via the same Web pages and interacted with the database in exactly the same manner. The laboratory group participated in the experiment in a room at the Department of Psychology of the University of Fribourg, where a computer was provided which had access to the database over the intranet of the university. An experimenter was present during the entire experimental session. The Web group accessed the experiment via the Web through a link provided in a post-registration e-mail.

Apparatus.

A Pentium I (133 MHz, 32Mb RAM) with a LAMP (Linux Apache MySQL PHP) environment was installed. Linux Mandrake 6.2 was used as an operating system to host the experiment for both the laboratory and the Web versions. MySQL 3.23.41 served as database, and PHP 4.0.7 generated the necessary Web pages. An Apache Web Server 1.3.20 connected the components to the Web. Note that LAMP provides a server-sided solution that needs only a browser on the client's (i.e., the participant's) side. No further installation of software is necessary; therefore, compatibility problems are rare. The use of a database to record the participant's clicks is an elegant method of data collection, because easy data access is assured. The participants were randomly assigned to the versions through a PHP query after registration.

The list version was generated via a list of all the questions used in the experiment. (See Table 9 for details on the number of questions and number of words relevant for the text search in the keyword version for each task.) In the keyword version, a full text search through the database of possible questions was used for information gathering.

Table 9. Number of Questions and Words for the Decision-making Task

Task	Questions	Words
Post	39	390
Ticks	89	932
Mechanics	106	1228

All actions taken during the experiment were recorded and written to the database with a timestamp (current time of an event that is recorded by a computer) for each individual step.

6.2.2 Results

Dropouts.

Analysis of dropouts is an important tool in online research. In the current experiment, the laboratory and Web conditions yielded different dropout behaviors. A dropout in this experiment was defined as someone who had registered (by submitting an e-mail address and receiving a password) and who did not start the experiment, had started but not finished all the tasks of the experiment, or just lurked through the experiment without clicking on any question. Forty-four (36.44%) participants dropped out of the Web sample; significantly more dropped out of the keyword version (30) in comparison with the list version (14; $\chi^2(1,\ N = 44) = 5.82$, $p < .016$). By contrast, the laboratory sample had no dropouts at all.

Clicks.

In this study, a click was defined as the clicking on a hyperlink in order to obtain information about a question of interest. This measure was used for both the list and the keyword versions. Clicks were analyzed using an ANOVA with location (laboratory vs. Web) and version (list vs. keyword) as between-subjects factors. Task (ticks vs. mechanics) was varied as a within-subjects factor. Task yielded no significant difference—the ticks and mechanics tasks resulted in comparable amounts of information. Location was significant $F(1{,}74) = 14.66$, $p < .001$, indicating that more items were clicked on in the laboratory condition ($M = 12.03$, $SD = .87$) than in the Web condition ($M = 7.91$, $SD = .64$; see Figure 15). Version was also significant $F(1{,}74) = 4.62$, $p < .035$. More items were clicked on in the list ($M = 11.13$, $SD = .78$) than in the keyword condition ($M = 8.81$, $SD = .74$).

In addition, a significant location X version interaction was found for the laboratory condition $F(1{,}74) = 6.88$, $p < .011$. A post hoc analysis of the laboratory condition revealed a significant difference between versions $t(52) = 3.01$, $p < .004$. More items were accessed in the list than in the keyword version. No effect was found for the Web condition.

In the keyword version, no differences between laboratory and Web were found. Methodological problems may account for this. For the keyword version, the number of dropouts was larger on the

Web than in the laboratory. It is important that a significant difference in the number of keywords entered was found between the two locations ($M_{laboratory}$ = 7.50, $SD_{laboratory}$ = 5.68 versus M_{Web} = 4.41, SD_{Web} = 3.70; $t(94) = 3.03$, $p < .003$. Although the participants in the laboratory sample entered nearly twice as many keywords to gain information, there was no difference in the number of clicks (see Figure 15).

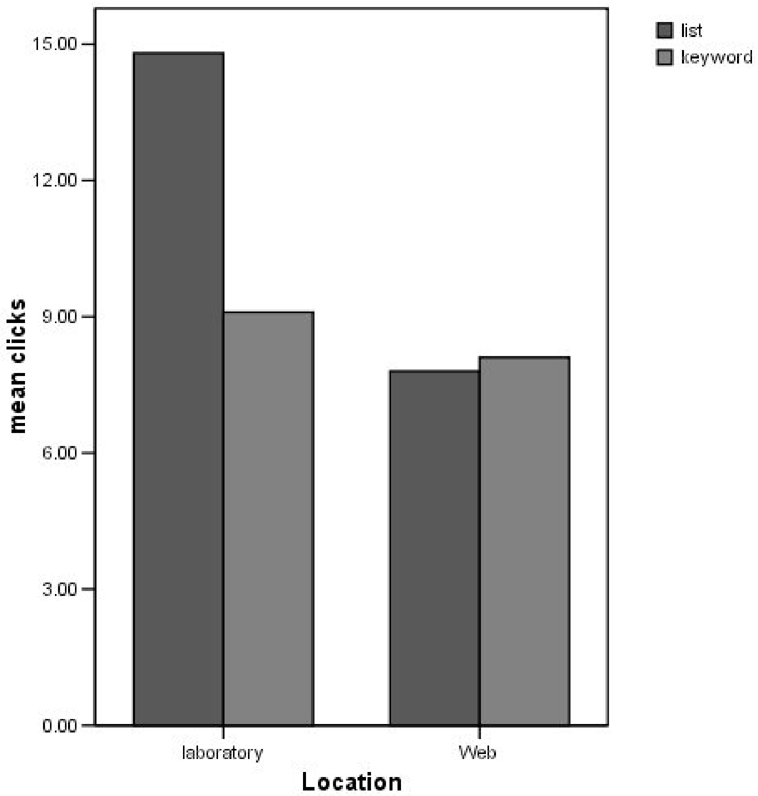


Fig. 15. Mean number of clicks (Experiment 1)

Categories.

On the basis of Huber and Macho (2001), six categories of information were formed for this analysis: *probability, new alternative, control, worst-case plan, situation, and consequences* (see Table 10 for descriptive statistics). An ANOVA with location (laboratory vs. Web), version (list vs. keyword), and category as between subjects factors was conducted. Clicks served as a dependent variable. A significant main effect was found for categories $F(1{,}900) = 98.83$, $p < .001$. A post hoc Scheffé test revealed that the participants used the categories *situation* and *consequences* more often than the other categories (see Table 10). These results are in accordance with the findings of Huber et al. (1997; Huber et al., 2001), confirming that *probability* information is of less interest than *situational* and *consequential* information.

A second analysis of categories was performed in order to obtain a measure of chosen categories unaffected by the number of information items examined. The dependent variable was the propor-

Table 10. Descriptive statistics for clicks (Experiment 1)

		Location			
		Laboratory		Web	
Version	Category	*M*	*SD*	*M*	*SD*
List	Probability	1.61	1.31	0.57	0.87
	New alternative	0.68	0.90	0.24	0.43
	Control	1.46	1.67	0.76	0.83
	Worst Case	0.86	0.85	0.57	0.65
	Situation	5.96	2.73	3.22	2.24
	Consequences	4.50	2.63	2.22	1.80
Keyword	Probability	0.93	1.27	0.77	0.98
	New alternative	0.26	0.53	0.03	0.18
	Control	0.56	1.05	0.52	1.05
	Worst Case	0.56	0.64	0.40	0.69
	Situation	4.00	3.34	3.26	3.89
	Consequences	3.15	2.63	3.26	3.27

tion of information searched for in each category by the participants. χ^2 tests on the difference between locations were performed. No significant difference could be found in any of the categories.

Time.

Each individual click in the experiment was automatically attached to a timestamp. This procedure makes it possible to calculate completion time for the tasks. Time was analyzed by an ANOVA with location (laboratory vs. Web) and version (click vs. search) as between-subjects factors. Location was significant $F(1{,}153) = 37.03$, $p < .001$, indicating that more time was used to complete the task in the laboratory ($M = 367.3$, $SD = 22.7$ sec) than on the Web ($M = 192.9$, $SD = 17.4$ sec). No effect was found for version.

6.2.3 Discussion

These results suggest that the location where the experiment was conducted had an effect on the number of items considered relevant. The participants in the Web condition did not use as much information as did those in the laboratory. This finding contradicts previous comparison studies that showed no difference between the laboratory and the Web (Buchanan & Smith, 1999; Epstein et al., 2001; Krantz et al., 1997; Krantz & Dalal, 2000; Mehta & Sivada, 1995). However, this effect was limited to the list version. The keyword version showed a ceiling effect on number of information items looked at, with certain hints of methodological problems in information acquisition. The differences observed in Experiment 1 can be interpreted as either an influence of measurement between Web and laboratory conditions or merely a difference between Web and laboratory participant samples. In Experiment 1, the influences of these two factors cannot be separated. This problem is inherent to most validity studies using Web methodology. The fact that in a study a difference between Web and laboratory conditions was revealed does not prove that there is a reliable difference due to administration conditions if the samples are different. A second experiment was conducted to overcome this confound.

6.3 Experiment 2 – The sampling problem

The goal of Experiment 2 was to solve the problem of different samples. This was done by assigning participants from a single population randomly to the Web and laboratory conditions (see Shavit et al., 2001). The experimental design and the decision-making tasks from Experiment 1 were used again.

6.3.1 Method

Participants.

The participants were recruited via an announcement at the University of Fribourg. Eighty students volunteered and were randomly allocated to either the laboratory or the Web condition.

An e-mail was sent to 40 students requesting their participation in the Web condition. Of these, 32 students (19 female, 13 male) with $M_{Age} = 23.75$ ($SD_{Age} = 4.64$) responded and completed the experiment. This represents a dropout rate of 20% for the Web condition (list version, 3; keyword version, 5; *n.s.*).

In the laboratory condition, several appointment times were proposed to the 40 students (26 female, 14 male). Every participant completed the experiment. The participants in this sample had $M_{Age} = 24.80$ ($SD_{Age} = 4.54$).

Each participant received course credit for participation.

Design, Apparatus, and Procedure.

The method, tasks, design, and apparatus used in Experiment 1 were used in this experiment as well.

6.3.2 Results

Clicks.

An ANOVA with location (laboratory vs. Web) and version (list vs. keyword) as between-subjects factors was conducted. Task (ticks vs. mechanics) was used as a within-subjects factor. Task yielded no significant difference; the ticks task and the mechanics task resulted in comparable interest in information. Location was significant $F(1,252) = 198.06$, $p < .001$, indicating that more items were clicked on in the laboratory condition ($M = 6.27$, $SD = .21$) than in the Web condition ($M = 2.06$, $SD = .22$; see Figure 16). Version was also significant $F(1,252) = 149.62$, $p < .001$. More items were clicked on in the list ($M = 5.99$, $SD = .10$) than in the keyword ($M = 2.34$, $SD = .23$) condition. In addition, a significant interaction between location and version $F(1,252) = 239.99$, $p < .001$ was found. A post hoc analysis of the laboratory condition revealed a significant difference between versions $t(199) = 19.81$, $p < .001$. More items were accessed in the list than in the keyword version. No effect was found for the Web condition.

An analysis of entered keywords in the two locations demonstrated a larger number of keywords entered in the laboratory condition ($M = 5.68$, $SD = 3.85$) in comparison with the Web condition ($M = 3.27$, $SD = 2.88$). This significant difference $t(118) = 1.98$, $p < .05$ emerged although there was no difference in the number of clicks.

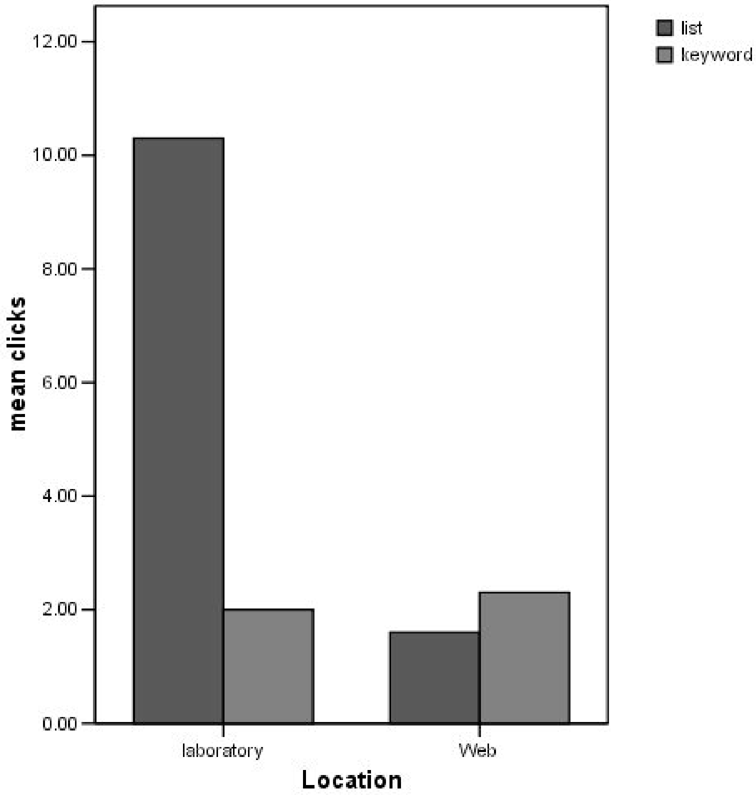


Fig. 16. Mean number of clicks (Experiment 2)

Categories.

An ANOVA with location (laboratory vs. Web), version (list vs. keyword), and categories (probability vs. new alternative vs. control vs. worst-case plan vs. situation vs. consequences) as between-subjects factors was also conducted. A significant main effect was found for categories $F(1{,}852) = 109.91$, $p < .001$.

A post hoc Scheffé test resulted in significant, more frequent use of the situation and consequences categories in comparison with the other categories (see Table 11). These results are in accordance with the findings of Huber et al. (1997; Huber et al., 2001), demonstrating that probability information is of less interest than situational and consequential information.

A second analysis of categories was performed in order to obtain a measure of chosen categories unaffected by the number of information items looked at. The dependent variable was the proportion of information searched for in each category by the participants. Note that this analysis works with a proportion of hits in a certain category and not with the absolute number of hits. This assumption enables the calculation despite the small number of clicks in three of the four conditions (see Figure 16). χ^2 tests on the difference between locations were performed. In no category a significant difference was found.

Table 11. Descriptive statistics for clicks (Experiment 2)

		Location			
		Laboratory		Web	
Version	Category	*M*	*SD*	*M*	*SD*
List	Probability	1.24	1.26	0.41	0.61
	New alternative	0.42	0.50	0.26	0.45
	Control	0.84	0.85	0.68	0.81
	Worst Case	0.15	1.68	0.59	0.66
	Situation	3.87	2.63	3.15	2.23
	Consequences	3.51	1.65	2.18	1.71
Keyword	Probability	0.85	1.18	0.80	1.04
	New alternative	0.21	0.48	0.04	0.20
	Control	0.45	0.97	0.57	1.14
	Worst Case	0.48	0.62	0.45	0.73
	Situation	4.94	4.29	2.57	2.91
	Consequences	3.03	2.48	3.10	3.05

Time.

Time was analyzed by an ANOVA with location (laboratory vs. Web) and version (click vs. search) as between-subjects factors. Location was significant $F(1,67) = 27.99$, $p < .001$, indicating that more time was used to complete the tasks in the laboratory ($M = 292.25$, $SD = 20.41$ sec) than on the Web ($M = 144.23$, $SD = 19.13$ sec). No effect was found for version.

6.3.3 Discussion

Because the online population is steadily growing and consequently becoming more and more diverse, it is different from the student population. Not only do age, sex, and educational distributions differ, but, for the most part, economic well-being also varies. The argument that measuring different populations (laboratory and Web) leads to differences in results is a common one.

The purpose of Experiment 2 was to overcome the confounding factors between location and population sample that were present in Experiment 1. It was shown that the control of population does not change the general result of differences between locations. Therefore, the interpretation that the laboratory and Web conditions have a central effect on information acquisition appears to be appropriate.

6.4 General discussion of Web versus laboratory

The purpose of this study was to investigate differences in information acquisition behavior with data collected in the laboratory and on the Web. The main results of our experiments can be summarized as follows:

1. Differences emerged in the amount of information gathered between the Web and the laboratory conditions. These cannot be explained by sample differences. Using the same sample (students), the results of Experiment 1 were replicated in Experiment 2.
2. The participants in the list version of the AIS method searched for more information than did those in the keyword version (in the lab). On the Web, such differences were not found.

3. Different dropout behavior is found on the Web. Fewer participants dropped out of the experiment in the list version than in the keyword version. The lab condition showed no dropouts at all.
4. If we look at the distribution of question categories as a dependent variable, we see no differences (i.e., no interaction) between locations and versions. This result is also valid if we use the percentage of participants who asked at least one question in the respective category.

Why do participants behave differently in the laboratory and on the Web?

The presented literature on comparisons between laboratory and Web suggests that there should be no difference between the two research locations. There is a crucial difference between those studies and ours—namely, the response mode of the participants. In a pen and paper or Web-based questionnaire, participants can estimate the duration of the task and the number of answers requested. In an information search task, participants are expected to search as long as necessary for them to complete the task. No hint of what constitutes a good amount of search depth is given; participants decide on their own, and their decisions are different when the experiment is run on the Web.

In combination with this, a second critical factor is the presence of an experimenter in the laboratory. There are at least two possible explanations, which are not mutually exclusive: (1) The experimenter is an authority figure for the participant and, in his or her presence, the participant is more inclined to make an effort; and (2) the presence of the experimenter signals for the participant that the experiment is important; therefore, the participant is motivated to search for more information. Approximately twice the number of entered keywords were measured in the laboratory in comparison with the Web.

For future studies, a laboratory condition with an experimenter leaving the participant on his or her own during the experiment would be of interest.

Why do people with the list version search for more information than do those with the keyword version (in the laboratory)?

In the list version, a given list of questions is scanned by the participants, and questions of interest are chosen. In the keyword version, the task is as follows: Think of a keyword, write it into the appropriate field, scan through the results, and choose an information item of interest. Cognitive load and, therefore, effort made in completing a task are obviously higher for the keyword version. Assuming that participants want to keep cognitive effort low, this factor could account for the differences.

In addition, methodological problems seemed to arise in the keyword version. Using a common search engine on the Web enables one to find results for nearly every possible keyword. This is because the number of stored pages reaches the billions on these search engines. In our experiment, the search space was limited to the number of questions in the database (an average of 98). The participants had to learn which keywords yielded results and which did not. Search behavior was therefore influenced by the number of successful keywords. A hint of problems in the keyword version is also given by the number of dropouts, which was much larger than in the list version.

An improvement of the keyword version seems important. This could be accomplished by enlarging the database of questions and using smarter search devices than a full text search.

Why does the dropout behavior differ between locations and versions?

Dropout is a feature of Web experimenting that should be controlled and reduced (Reips, 2002). Nevertheless, a positive aspect of dropout is the insight into possible confounding of experimental conditions. In our experiment, we minimized the dropout rate by asking for personal information and giving information about incentives at the beginning of the experiment (Frick, Bächtiger, & Reips, 2001, for experimental manipulation of these two factors). Nevertheless, the two locations, laboratory (no dropouts) and Web (36% dropout rate in Experiment 1, 20% in Experiment 2), resulted in different dropout behaviors. Three possible explanations, which could be confounded, are suggested. The participants had problems in the keyword version because information search was not always immediately successful (see Reips, 2002a, for a discussion of the detection of motivational confounding). The presence of an experimenter in the laboratory could be a second reason: The participants were socially *trapped* and, therefore, forced to finish the experiment. It does not seem appropriate to leave an experiment when another person (e.g., the experimenter) is present. In the Web version other – not controllable – factors can have an influence, e.g., the participant is disturbed by another person.

For which dependent variable is the difference between laboratory and Web relevant?

If one is interested in the absolute amount of information gathered, there are certain differences between the two locations. Nevertheless, it is not clear from these two experiments which of the two locations is closer to reality.

Often, however, the researcher is interested not in the absolute amount of information, but in the relative amount of searching for different categories and changes in the relative frequency of categories in different experimental conditions. As our results show, the distribution of information search over categories is not affected by the laboratory—Web variation.

We investigated the difference between laboratory and Web experiments using a decision-making task as an example. However, our results are also relevant to other areas of psychological research in which information search behavior is of interest. Examples are the areas of complex problem solving, mental models, knowledge representation, and the evaluation of Web page content.

A scenario for the last point could be the following: The aim is to communicate information about a risky situation (e.g., breast cancer) on a Web page. An AIS experiment is run in order to locate types of information central to the participants. Finally, an analysis of these categories and a comparison with objective important information should reveal deficiencies in information gathering. With emphasis on this missing information, an improvement of information presentation should be possible.

To recapitulate, our experiments reveal that there is no simple answer to the question of whether or not a Web study is acceptable for information search tasks. Even if we find distinct differences in search behavior between the laboratory and the Web, the answer is highly dependent on the behavioral aspects in which the researcher is interested. A promising point is that the type of information demanded in a Web study seems to be comparable with that demanded in the laboratory study. Similar information is requested whether or not the experimenter is present.

7

WebDiP – Technical description and an experiment

This chapter[1] has two aims, the first of them is to give an extended description of WebDiP (Web DecisIon Processes) in order to demonstrate the technical capabilities of the system. The second aim is to highlight improvements from an earlier software version used in Chapter 6. Because in the earlier version especially dropout behavior yielded differences between the research locations, this measure and the amount of accessed information will be used as dependent variables for the comparison with WebDiP.

7.1 Introduction

With the rise of the Web the value of information and the corresponding search for it has grown quickly over the last 10 years. Search engines are the gatekeepers to information on the Internet providing fast and easy access to every imageable topic. The penetration of search engines into everyday life has even reached the language when the process of searching for information on the Internet is termed "to google" nowadays. This chapter introduces a research tool – WebDiP – that is based on the ideas of a search engine but adds capabilities for conducting experiments on information search in a controlled setting. The roots of the idea to construct such a program not only stem from the broad usage of search engines. Clearly various process tracing approaches introduced in Chapter 3 and critical points in using these (Chapter 3.5, p. 19) play an important role.

The starting points for the WebDiP project are the AIS framework (see Chapter 3.4, p. 16) as one cornerstone and the aim to automatically gather detailed information about the search process (used for example in the Mouselab system, Chapter 3.2, p. 13) as the other cornerstone. The aims of the program are the following: (1) a system that lets various researches run experiments with one WebDiP installation. (2) Web-based setup and simple administration of an experiment. (3) Export of data as well as experiments and tasks. (4) An easy to use interface for the participants which lets them concentrate on the task and not on the handling.

Based on these aims WebDiP was developed, tested and released on sourceforge.net[2] early in 2004. All this was done under the GNU General Public License (GPL). The GPL is intended to provide freedom to share and change software to make sure it is available to all interested users.

[1] A paper based on this chapter was published as Schulte-Mecklenbeck, M. & Neun, M. (in press). WebDiP - A tool for information search in decision making experiments on the WWW. *Behavior Research Methods.*

[2] an open source plattform: http://webdip.sourceforge.net

7.2 Technical description

The system design[3] is best viewed from two separate perspectives: (1) the experimenter's or administrator's perspective and (2) the participant's perspective. For (1) the following functions should be available: management of several parallel and independent experiments; e-mail verification for each participant (to overcome double entries); an easy user interface for experiment generation; the recording of a clickstream (i.e., succession of mouse clicks) from each participant; statistics showing an on-line status of an experiment; several export filters for data aggregation and further analysis in a common statistical software package. For the participant (2) the main objectives were a system that is accessible with a large number of browsers; an intuitive interface that needs only few instructions in usage terms and sufficient results for the information search in order to avoid a large number of dropouts (in Chapter 6 it was shown, that subjects reached a ceiling effect in their searches; they were sensitive to the response they got from the system, because their searches often showed empty results for their keywords).

7.2.1 Database design

This section describes all instances included in the WebDiP system. An instance can be understood as a part (or table) of the database. In the figurative sense each instance represents the parts of an AIS experiment[4].

ADMIN Each ADMINistrator[5] is defined through a login name, a password, an e-mail-address, a real name and an institution where he belongs. One *super user* is defined through the login name *root*, he has rights to create other ADMINs. Each EXPERIMENT ADMINistrator can own an arbitrary number of EXPERIMENTs. These two levels of adminstration provide each *root* ADMINistrator, e.g., a supervisor of 4 student groups, with rights to grant each group of EXPERIMENT ADMINs unique access to only "their" EXPERIMENT.

EXPERIMENT An EXPERIMENT is the complete framework needed to run a study. It contains the title, the general instructions, the EXPERIMENT's language (at the moment, German and English are implemented), the SEARCH MODEs, a warm up TASK and an arbitrary number of decision TASKs. Participants register for one specific EXPERIMENT and therefore *belong* (in database terms) to only one EXPERIMENT.

SEARCH MODE Every EXPERIMENT can make use of up to four different INFORMATION SEARCH MODEs. The SEARCH MODEs are *list* (a list of all INFORMATION items is shown to the participant, this mode was also used in Schulte-Mecklenbeck & Huber, 2003, see Chapter 6), *category* (the participant sees a list of categories first, after clicking on a category the containing INFORMATION is shown), *fulltext* (a Google-like search which shows a list of INFORMATIONs after entering of a keyword, see technical details in Chapter 7.2.2, p. 69) or *no search* (a condition where it is possible to present choices only, this condition could be used to generate simple questionnaires or classical decision making experiments with instructions and choices only). Every SEARCH MODE has a type, a title and an individual instruction.

PARTICIPANT PARTICIPANTs register for a specific EXPERIMENT. Each PARTICIPANT entry in the database contains a valid e-mail-address, the IP address (which is recorded automatically), the login date, the age and the gender of the PARTICIPANT. A PARTICIPANT

[3] The following description of WebDiP is based on Neun (2004) who has written a Master thesis at the Department of Informatics, University of Fribourg. Neun (2004) developed the WebDiP software in collaboration with the author of this thesis.

[4] an exception is the LOG (logfile), which solely can be accessed in WebDiP and has no representation in the AIS framework

[5] all instances that are part of the database design are written upper case to ease readability

will always be automatically assigned to one of the active SEARCH MODEs and uses this SEARCH MODE through the whole EXPERIMENT. For security reasons the PARTICIPANT entry also "knows" the actual TASK a PARTICIPANT is in. This prevents the PARTICIPANT from doing the same TASK twice by using, e.g., the browsers back button.

TASK A TASK represents one of an arbitrary number of decision TASKs. It contains its name and the TASK itself. A TASK can have an arbitrary number of choices. Usually these are at least two. Each TASK contains an INFORMATION database, in which the PARTICIPANT can search for INFORMATION using one of the SEARCH MODEs.

TASK CHOICES The TASK CHOICES consist of the choices the PARTICIPANT can make which are defined through the TASK text.

INFORMATION The INFORMATION is generated by the EXPERIMENT ADMIN before the task. The form of one INFORMATION item is normally of *question + answer*. INFORMATION is grouped into different categories (e.g., probability or consequences). For improving the search results, INFORMATION also contains the *soundex values* and the *stems* (see below) of the words in the question.

LOG The LOG records all actions of the PARTICIPANT while performing an EXPERIMENT. These are the time, the action itself (search, click, choose) and the value of the action.

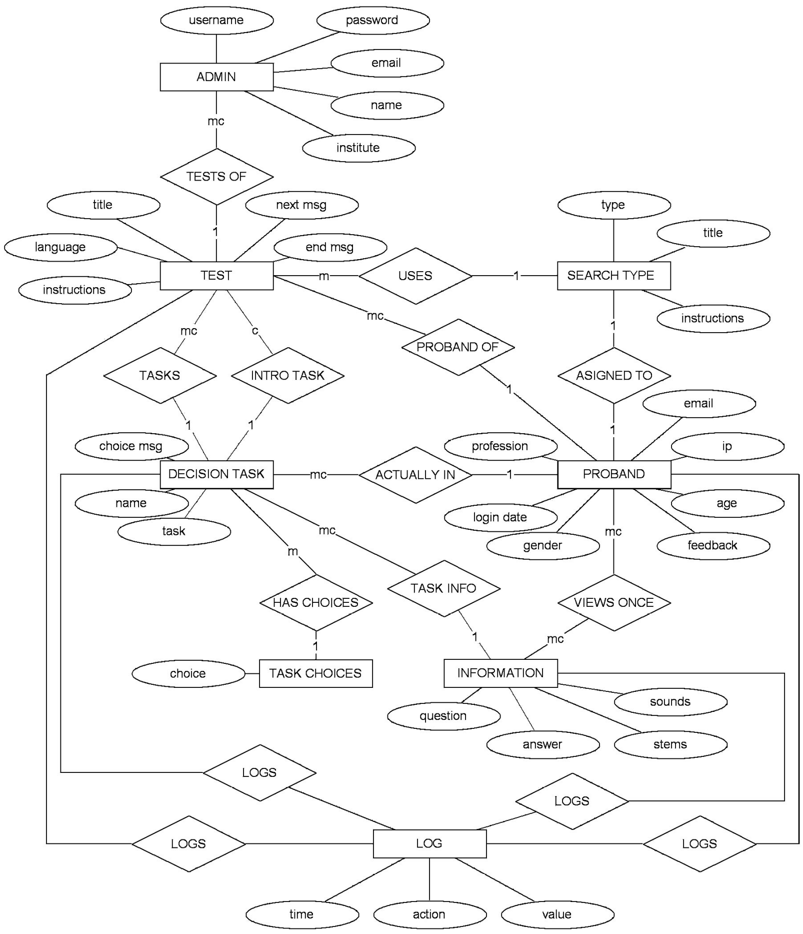
The above described instances and their relations are plotted in the *entity relationship model (EM)* of WebDiP in Figure 17. EMs are a graphical representation of entities (objects, i.e., ADMIN, EXPERIMENT ...) used in the creation process of a database[6]. The central importance of the EXPERIMENT (i.e. TEST), the TASK and the PARTICIPANT (i.e. PROBAND) is documented in this Figure, too. Additionally, the LOG's connections to all available entities of the system is clearly depicted. These connections enable the researcher to receive a detailed picture of the PARTICIPANT's moves during an experiment.

7.2.2 Information search

The development of comprehensive database search mechanisms was one of the central aims in the improvement process from the basic version of experiment 1. Because of the relative smallness of the INFORMATION a good search mechanism is important. Google for example has an indexed database of 4.285.199.774 Web pages (retrieved on June, 10th, 2004). This large database "guarantees" search results with nearly any keyword. Additionally, Google uses a sophisticated *Page ranking* (see http://www.google.com/technology/ for technical details, retrieved July 12th, 2004) mechanism which calculates all links to a Web-page and weighs each link according to the source Web-page's popularity. The first part of this recursive calculation can be compared to the common journal impact factor calculations. The popularity rating would add a higher rank to an important journal (e.g., Journal of Experimental Psychology) than to a less important Journal (e.g., Psychoanalytic Psychotherapy) providing a clearer judgment criterion. In comparison to Google, the INFORMATION tables for one TASK used in WebDiP have sizes of around 100 items. Additionally the TASKs used here are very specific, this is clearly a drawback. The search functions implemented in WebDiP are threefold:

1. a simple fulltext search is performed which compares the entered keyword of the PARTICIPANT with every word in the *questions* (in the INFORMATION table).
2. *Soundex* keys are generated of the questions. Soundex keys are based on Knuth (1977, as cited in Neun, 2004) and transfer words pronounced similarly into the same soundex key. They can thus be used to simplify searches where one knows the pronunciation but not the spelling. The

[6] Note that I have changed the names of the following entities in Figure 17 for easier comprehension: test=EXPERIMENT and proband=PARTICIPANT



mc=an arbitrary number including zero; m = an arbitrary number, but minimum 1

Fig. 17. Entity relationship model of WebDiP.

soundex function returns a string of 4 characters starting with a letter. The substitution code is based on the following rules. The first letter of the word is always retained. The rest of the word is compressed to a three digit code based on the following coding scheme:

```
A E I O U Y H W     not coded
B F P V             coded as 1
C G J K Q S X Z     coded as 2
D T                 coded as 3
L                   coded as 4
M N                 coded as 5
R                   coded as 6
```

Consonants after the initial letter are coded in the order they occur:

```
HOLMES = H-452
ADOMOMI = A-355
```

The code always uses an initial letter plus three digits. Further consonants in long names are ignored:

```
VONDERLEHR = V-536
```

Zeros are used to pad out shorter names:

```
BALL = B-400
SHAW = S-000
```

As are adjacent consonants from the same code group:

```
JACKSON = J-250
```

3. A stemming algorithm was developed which searches for words with the same meaning. This is accomplished by automatically removing suffixes from the word stem. Terms with a common stem will usually have similar meanings, for example:

```
CONNECT
CONNECTED
CONNECTING
CONNECTION
CONNECTIONS
```

In this example removing of the various suffixes -ED, -ING, -ION, IONS leaves the single term CONNECT. The stripping process will reduce the total number of terms in the database as well as the size and complexity of the data in the system.

The three steps described above considerably improved the system's performance when participants searched through the INFORMATION table with keywords.

7.2.3 User interface examples

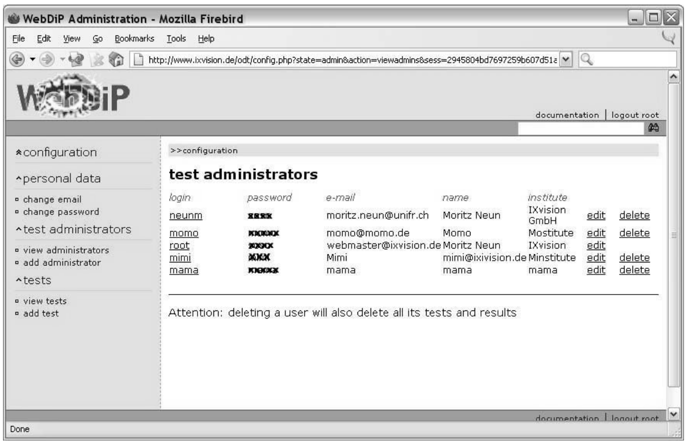


Fig. 18. Experimenter management

This section wants to show some parts of the user interface in WebDiP and explain their functions. For a detailed overview the online release of the system should be consulted at: http://webdip.sourceforge.net.

Management of Experimenters (see Figure 18) can be done by the ADMIN in the management window. The ROOT (superuser) user can create new EXPERIMENT ADMINs (other experimenters), change their properties or login as one of them to be able to easily check for errors or support them.

The *experiment settings* window enables the EXPERIMENT ADMIN to create or import new EXPERIMENTs, to modify already existing ones or delete old EXPERIMENTs. Each new EXPERIMENT receives a link (see Figure 19) which is also used to invite participants to the experiment. Note that each link to an experiment contains a 32 character code, which ensures that participants can not access other conditions in an experiment through changing digits or characters in a link. This method guarantees the integrity of each experimental condition and therefore protects the researcher from problems addressed in, e.g., Reips (2002).

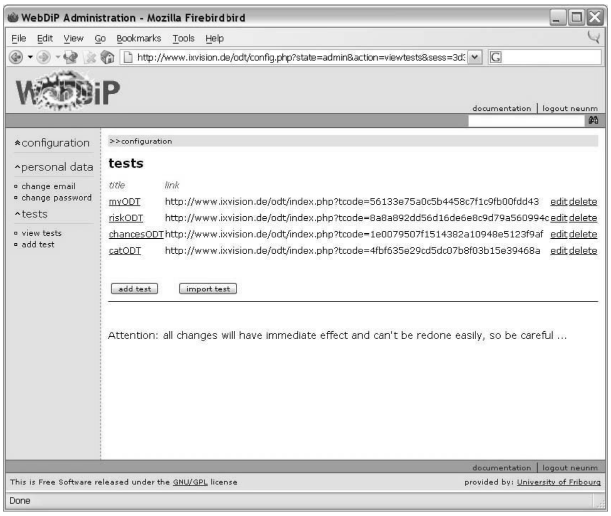


Fig. 19. Test management

On the next level, the EXPERIMENT's options can be set (see Figure 20). Additional to the title, the language and the basic instructions for the participants the SEARCH TYPES can be configured here. Four SEARCH TYPES are available which are described in Chapter 7.2.1, p. 68. The STARTTASK (which is always presented as the first task) as well as the MAIN TASKS are additionally defined here.

The options for a single TASK are defined within edit task (see Figure 21). Each task must receive a NAME (which is also shown on the participants screen during an experiment) and a TASK TEXT which can be edited with common HTML commands for better readability. The CHOICE MESSAGE contains the actual question which must be answered by the participant. Connected to the CHOICE MESSAGE are an arbitrary number of CHOICES which in the current experiments do not exceed two. Finally, a list of INFORMATION CATEGORIES with the number of containing INFORMATION ITEMS is presented.

On the lowest level, the INFORMATION is entered (see Figure 22). Each INFORMATION contains a *question* and *answer*. Additionally, a CATEGORY for each INFORMATION has to be entered. Previously entered categories can be accessed through a dropdown list in order to guarantee consistent entering of information.

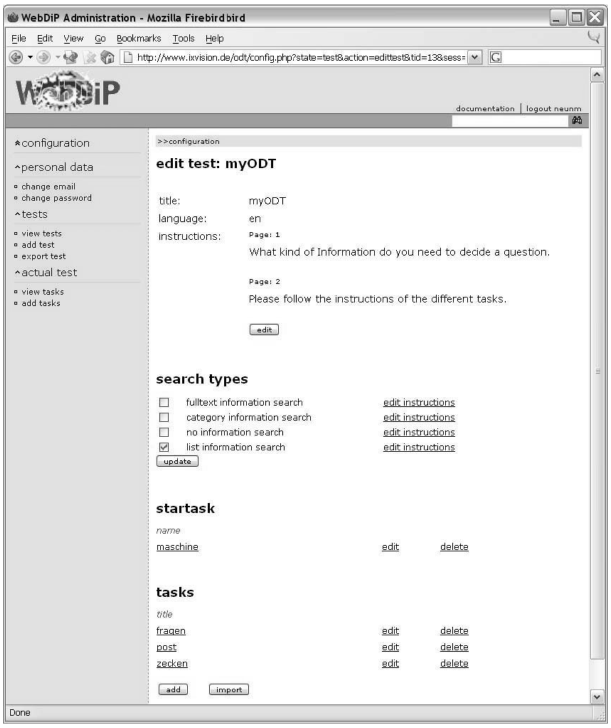


Fig. 20. Configuration of an experiment

On each of the above levels it is possible either to *export* current data or to *import* already saved ones. This feature enables easy distribution of INFORMATION, TASKs or whole EXPERIMENTs. In the statistic section, the EXPERIMENT ADMIN is able to view bar charts of CATEGORIES with current access information per TASK. An easy overview of what is going on during an experiment is possible through this feature. Additionally, various statistics can be downloaded in this section, a whole LOG-file is available as well as statistics on, e.g., PARTICIPANTS per TASK, CHOICES per TASK or accessed INFORMATION.

Moving on from this description of technical capabilities of WebDiP, the following experiment wants to demonstrate the suspected improvements with a comparison of WebDiP to experiment 1 and 2.

7.3 Experiment 3 – Comparison between study 1 and WebDiP

Study one showed several methodological shortcomings which should be overcome through a redesign and improvement of the used software. The aim of the new software was to improve the keyword version because this condition is understood to be the most *common* one for subjects (comparable to a Google search). Additionally, this version seems to be particularly useful in the planned experimental design due to the total lack of hints (on categories of information) for the participant.

In Chapter 6 (p. 53), certain shortcomings like, e.g., a low number of clicks in the keyword version, are addressed. To demonstrate the technical enhancements of WebDiP in comparison to the older version the following experiment was run. Data from experiment 1 (Post Office Task)

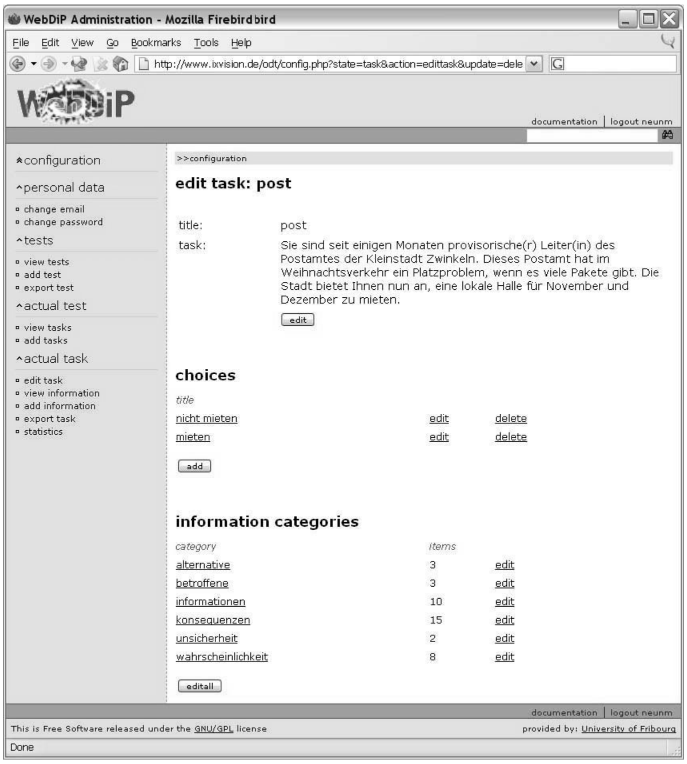


Fig. 21. Task Options

are compared with data from the present study (with the same task) done with WebDiP. For this analysis, participants from the Web-condition from study one are put into one group. From this group only the keyword version is analyzed – this group has the same requirements and can therefore be compared to the WebDiP sample.

7.3.1 Method

Participants.

Participants for the WebDiP study were recruited through a panel of a German market researcher (forschungswerk.de) and therefore represent a quite mixed sample (students replied as well as workers or employed). E-mails were sent to 500 randomly selected addresses of the panel with a short description of the purpose of the experiment, a link for registration and a description on a lottery with prizes[7]. From the 500 recipients of e-mails, 223 (44.6 %) registered for participation. Of these, 147 (65.9 %) finished the experiment (this means an overall participation rate of 29.4 %), 93 participants were male with $M_{Age} = 38.96$ ($SD_{Age} = 11.3$) and 54 participants were female with $M_{Age} = 32.52$ ($SD_{Age} = 10.8$). Data were collected between May 3rd, 2004 and May 10th, 2004. The combination of the Web groups from experiment 1 and 2 results in 108 participants (51 female, 57 male) with $M_{Age} = 26.24$ ($SD_{Age} = 4.32$). Only participants in the keyword version of both experiments were used.

[7] For a detailed description of the registration process see the Method section in Chapter 8, p. 77

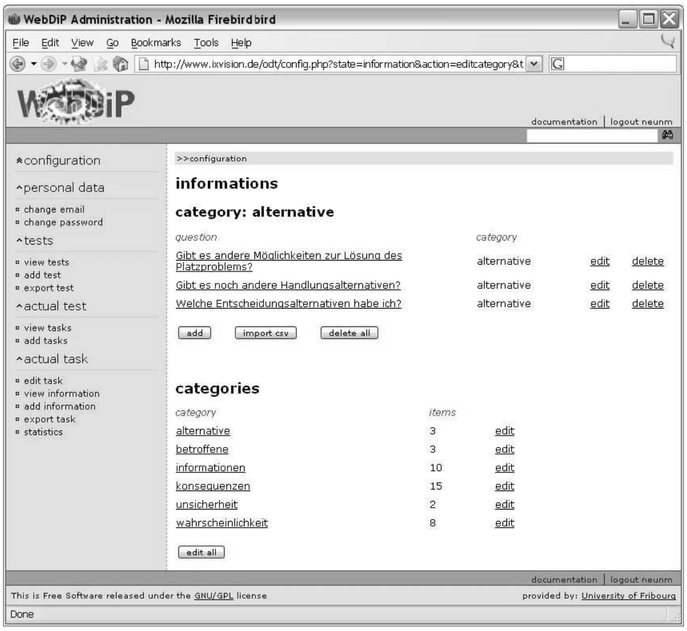


Fig. 22. Information Options

Design.

The comparison of WebDiP and the method from experiment 1 and 2 was of interest. Therefore, one independent variables was used in this experiment: *tool* (WebDiP X keyword version study 1). The same task was used in both studies (Post Office Task). As the dependent measure, the amount of clicks in general and per category were analyzed.

Apparatus.

A[8] Pentium 3 (800 MHz, 256Mb RAM) with a LAMP (Linux Apache MySQL PHP) environment was installed. Linux Mandrake 9.0 was used as an operating system to host the experiment. MySQL 4.0.15 served as database, and PHP 4.3.3 generated the necessary Web pages. An Apache Web Server 2.0 connected the components to the Web. WebDiP in version 1.41 provided the experimental setup for confirming participation, offering a search environment and storing the data.

7.3.2 Results

Dropouts.

The number of people who registered for the experiment (showed some basic interest) and did not finish it was used as a definition of dropout. In experiment 1 and 2, 36.44 % dropedout, in the WebDiP study this number was 34.1 % (n.s.). On this measure, no increase in participants finishing the experiment was reached because of the modifications.

[8] This description deals only with the WebDiP experiment. Details about Apparatus of Experiment 1 (6, p. 53) can be found in Chapter 6.2.1, p. 57

Clicks.

The Click definition of experiment 1 and 2 is also used here. The basic observational unit in WebDiP therefore remains a click on a hyperlink by a participant in order to obtain information. Inspecting the mean number of clicks between the categories of WebDiP and Study one (see Table 12) it becomes clear that an increase of clicks in all but one category (*situation*) can be observed in WebDiP. This impression can also be confirmed when the means of clicks between the two versions are compared. WebDiP users clicked on 2.44 ($SD = 0.15$) hyperlinks whereas participants in experiment 1 and 2 only clicked on 1.75 ($SD = 0.21$) an increase of nearly one overall click $t(378) = 2.05$, $p < .041$.

Table 12. Mean number of clicks between Studies

		WebDiP		Study 1	
Version	Category[a]	*M*	*SD*	*M*	*SD*
Keyword	Probability	1.87	1.44	0.49	0.52
	New alternative	1.46	0.59	0.25	0.44
	Control	1.67	1.37	0.72	0.41
	Worst Case	1.68	0.99	0.58	0.53
	Situation	2.93	2.78	3.19	2.14
	Consequences	5.17	5.38	2.20	1.79

[a] In order to be able to compare the data from experiment 1 and 2 with the WebDiP data the same coding schema is used.

7.3.3 Discussion of the comparison

It was shown, that with WebDiP, an increase in clicked on information can be found. Compared to Schulte-Mecklenbeck and Huber (2003)'s experiment 1 and 2, mean clicks increased with the same task but an improved environment. In an online experiment, a more straightforward observation of dropout behavior is normally possible than in a laboratory situation because of the greater distance between the participant and the experimenter resulting from the medium. Lesser complaints about problems with the environment were received with WebDiP in comparison to experiment 1 and 2 – this is understood as evidence for an increase in quality of the testing environment.

One difference between the two experiments remains – the sample. The sample of the WebDiP study was considerably older than that of experiment 1 and 2. Additionally the e-mail addresses were taken from a panel of a market researchers. All people on this panel had already taken part in a survey and acknowledged their interest in other surveys or studies. One striking fact was the reduction in collection time of the data. The two months from study one were reduced to 10 days in the WebDiP study. Given this much shorter time for testing, the usage of panels for further research questions seems reasonable.

The main aim of this experiment was to demonstrate the increased usability and usefulness of WebDiP in comparison with the old testing environment from experiment 1 and 2. Because of the gathered data, it is justifiable to use WebDiP for the final study with a Web sample only.

8

Study 2 – The domain issue

8.1 Introduction

In Chapter 5, p. 41, it was suggested that there is a need for a differentiation between a *decision mode* (the state a decision maker is in when performing a decision) and a *decision domain* (the content a decision is about or the area it is from). The literature reviewed in the above chapter handles the two termini interchangeably despite fundamental differences between them. *Decision modes* have to be induced (comparable to emotions) or the decision maker has to imagine being in a certain decision "mood" (e.g., rational or making a "gut" decision). *Decision domains* define aspects necessary for a decision through the content they are about (e.g., things to buy or drugs to take).

The aim of the fourth experiment is to test whether domain differences described in Chapter 5 can be found in an information search experiment (decision modes are not taken into account in this experiment because the still open question of how a decision mode can be induced). If the assumption holds, that a decision domain elicits different decision behaviors will be tested in this last experiment. Of special interest is the confounding between the effect of *decision domain* and *decision task* in the current literature. A task is always taken as representative for a domain, i.e., every domain (from the reviewed literature) corresponds to *one* task the researchers judges as fitting. A task about a drug to take is judged as having a good fit in the medical domain, whereas a task that deals with buying something clearly is judged as being representative for the business domain. Whether this assumption of "representativeness" holds using an information search approach will be tested in an online experiment using two (instead of one) tasks per domain.

The following components, introduced earlier in this thesis, are used in this chapter: The results from the survey on domains run in Chapter 5, p. 41 define the domains for the subsequent tasks. Three domains are chosen as target once: *business, medicine* and *law*. For each of them two tasks are generated which are completely parallelized in terms of *structure, outcomes and probabilities*. A second factor used in this experiment checks whether differences in task texts (in terms of amount of presented information) change the choice or search behavior of the participants. Therefore three different *cue versions* were created with the levels low-, medium- and high-amount of information about the task (see a detailed description below). WebDiP, which was introduced in Chapter 7, p. 67, serves as the experimental tool for this endeavor.

The following research questions should be answered:

(1) Do two tasks ascribed to one domain (by the researcher) result in comparable choices and overlapping interest in information? Given the results to this question: Can two tasks be combined to one domain?

(2) Does additional information in task texts (concerning hints on the structure of a task) change information search behavior or choices?

(3) Given the same structure and values in a decision task, does content change the information search behavior?

8.2 Experiment 4 – Domain differences and WebDiP

8.2.1 Method

Participants.

Invitation e-mails (see Appendix, p. 121) were sent to two panel lists. The first panel is run by the University of Erlangen-Nürnberg, Economic- and Social-Psychology department and is located at http://www.wisopanel.uni-erlangen.de/. The second panel is run by Jörg Hartig at http://www.joerghartig.de/. The panels are different concerning their target population in the following sense: the Social-Psychology panel has questionnaires and marketing research as the main topics. Experiments are also run within this panel but to a lesser extent. The "Hartig panel" clearly aims at participants interested in psychological research in general. The combination of the two should result in a relatively widespread sample in terms of the common parameters age, sex, income and interests.

503 registrations for the experiment were recorded between July 30th and August 13t h, 2004. From the 503 registrations 368 (73%) finished the experiment[1]. Of these participants, 226 were male with $M_{Age} = 35.41$ ($SD_{Age} = 11.80$), 142 were female $M_{Age} = 31.16$ ($SD_{Age} = 9.62$).

Apparatus.

A Pentium 3 (800 MHz, 512Mb RAM) with a LAMP (Linux Apache MySQL PHP) environment was installed. Linux Mandrake 9.1 was used as an operating system to host the experiment. MySQL 4.0.15 served as database, and PHP 4.3.3 generated the necessary Web pages. An Apache Web Server 2.0 connected the components to the Web. WebDiP in version 1.43 provided the experimental setup. This included participant management, the search environment and storage of the click data (see details about WebDiP in Chapter 7).

Design.

Two independent variables were examined: *domain* (business vs. law vs. medical) and *amount of cues* (small vs. medium vs. large). The independent variable *domain* was varied within participants, whereas *amount of cues* was varied between participants (see Figure 23). To be able to check for order effects a Latin Square was used for shuffling the tasks.

Tasks

The top nominations of the resulting *decision domains* from survey 2 (Chapter 5.3.1, p. 50) were used for the different tasks. From these domains three are taken into further consideration: business, law and medical. The personal domain is not used because of the problem to find factors like "effect of action *a*" within this concept.

For each *domain* two tasks are generated (resulting in $3 * 2 = 6$ tasks). These are for the BUS domain the *drilling* and the *pipe* task: the *drilling* task is about the choice between two drilling

[1] This equals a dropout of 27% which is a considerable improvement to study 1 which had a dropout of approx. 35%.

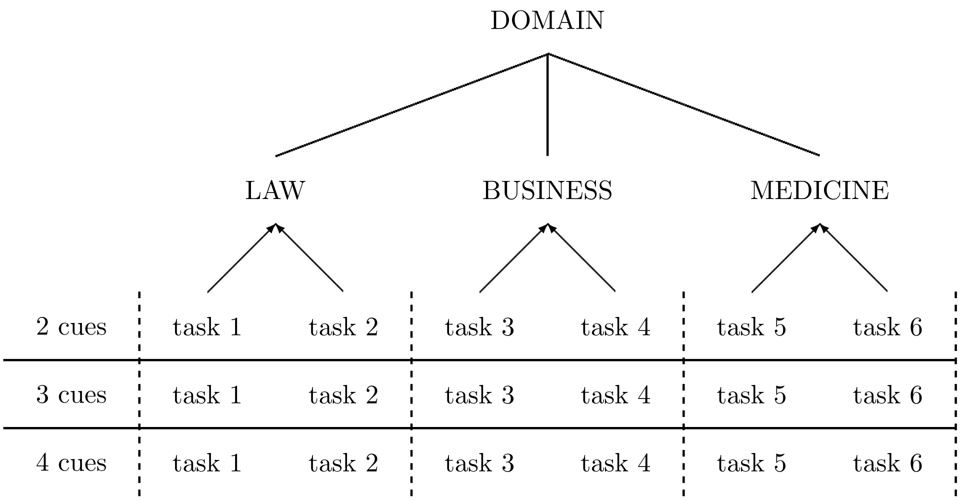


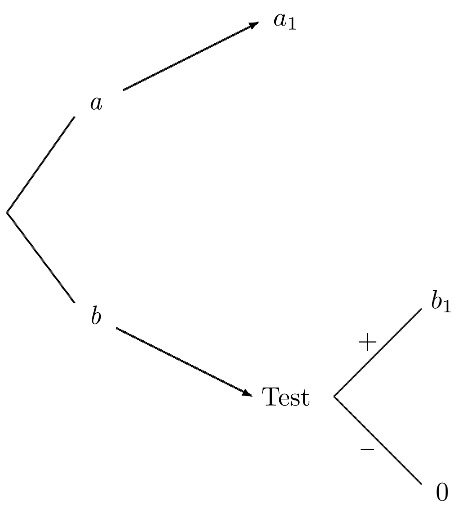
Fig. 23. Experimental design

systems for a company, which wants to drill a tunnel into a mountain. The *pipe* task is about the choice between two companies mending a broken pipe in the decision maker's home. In the MED domain the *drug* and the *operation* task are used. The *drug* tasks deals with the choice between two drugs for a lethal disease. The *operation* task is the choice between two operations against a gastric tumor. Finally the LAW domain includes the *jail* and the *rehabilitation* task. In the *jail* task the decision maker has to choose between two jails for a prisoner. In the *rehabilitation* task two rehabilitation programs for a prisoner are the choices. The complete task texts in German can be found in the Appendix A.5, p. 121.

The basic structure of all tasks follows the same schema (see Figure 24): There are two alternatives a and b to choose from. Alternative a is always connected to a subsequent action a_1. This action (a_1) has to be conducted when a is chosen (e.g., when drug a is taken an additional drug a_1 has to be taken). With the other alternative (b) a second stage is introduced. Whenever action b is chosen there is an external *test* which is run without influence of the decision maker. Connected to this *test* there is a chance that action b_1 has to be taken but also a chance that nothing additional (in case alternative b is chosen) has to be done. An example for this situation from the jail task is: given prison Tierau is selected (alternative b) a probation test (*test*) would determine whether the prisoner wants to escape and should therefore be held in high security area (*b1*) or whether no additional measures are necessary).

The following probabilities were assigned to the components[2]: a and b have a likelihood of success of 61% respectively. The two subsequent actions a_1 and b_1 have a high likelihood of success (if combined with a or b) of 91%. The tests general accuracy is high (92%), false positive (incorrectly reporting that something has been found) is 7.6% and false negative (reporting that nothing was found, when actually there was a result) is 0.4%.

[2] Every probability or other digit (e.g., costs of an option) is kept constant in all tasks



a and b: the alternatives to choose from
a_1 and b_1: subsequent actions, i.e. the outcome of a choice
Test: an external event choosing between the two outcomes *b1* and *0*, that can not be influenced by the decision maker

Fig. 24. General structure of the decision tasks

An additional second factor to domain – *cues* – is introduced. It has three levels: low, medium and high amount of information in the task text. The just described tasks in their basic version equal the *2 cues* condition – the two alternatives are understood as the two cues here. For the *3 cues* condition a sentence is added to each of the task texts which provides a hint on additional *side effects*. An example taken from the *operation* task (MED domain) is *internal bleeding* (as the side effect) that could occur after the operation. In the *4 cues* condition the whole task structure (see Figure 24) is presented to the participant with the two alternatives from the 2 cues condition and the two additional actions a1 and b1. An example for this condition, taken from the drilling task, is the hint on the existence of additional accessory units Varim (a1) and Sertec (b1) for the two machines (see Table 13, p. 81 for an overview of the two alternatives, the additional actions and the test for all six tasks).

For each task elaborated pre-testing resulted in a large number of questions and answers for the search process (see Appendix A.5.1, p. 124 for a list of all questions for the six tasks). All the central variables (e.g., a, b_1 ...) were parallelized. In order to keep the conditions as balanced as possible 99 questions and answers were selected (the selection criterion was whether is was possible to find a question and answer for all six tasks with corresponding meaning) as the data base for the experiment.

An important feature of the two alternatives is the precondition to be as comparable as possible. Equality (1) guarantees a maximum time for the search process and (2) opens the possibility to ascribe resulting effects to the content of a task or domain. The achievement of this equality is relatively easy within one task (some small changes between a and b were introduced within each task to keep the tasks interesting for participants) however, between many tasks certain problems arise: An example is the probability of a disease in a population (the incidence) – in the medical task 5% incidence is already quite high. In comparison, a 5% chance of getting an contract in the business domain is considerably low.

The central components of each task can be inspected in Table 13. In this table the two alternatives (a,b) and the connected additional actions are listed, too. The complete task-texts for the three cue levels can be found in the Appendix (p. 121 et seqq.).

Table 13. Transition of task-components

Domain	BUSINESS	MEDICINE	LAW
Task	drilling	drugs	jail
a	Brosch	Spinox	Hinterheim
b	Holcim	Lofa	Tierau
a_1	Varim	Tremol	solitary
b_1	Sertec	Rigolin	high security area
Test	sensor	blood test	probation test
Task	pipe	operation	rehabilitation
a	Bucher	Inselspital	Lenkheim
b	Steiner	Südklinik	Seewald
a_1	complete rehabilitation	endoscopy	rehabilitation
b_1	refurbishment	gastroscopy	community service
Test	Magnetometer	blood test	personality test

The position (POS) of the tasks in the experiment is controlled by a Latin-square design. The variation of task position is programmed into WebDiP and executed automatically while a new experiment is generated.

Procedure

Participants in this experiment were recruited via two panels. (see the *Participants* section above). Each of the participants received an e-mail with an invitation to participate and a description of the subsequently played lottery as a teaser. After clicking on a link in the invitation e-mail participants were randomly assigned to one *cue condition*. The registration process demanded the participant's e-mail address, gender, age and profession. After this step a second e-mail with a link to the corresponding condition was sent out. Participants clicked on the new link in the second e-mail and were redirected to an instructions page explaining the basic functionalities of the experimental setup. This step was followed by a warmup task (the *post* task (see Table 8, p. 57) which has been used as a warmup task in several other experiment's because it is especially simple in terms of structure and lets participants concentrate on the functionalities of the program). The warmup task was followed by three main tasks each from one of the domains MED, LAW and BUS. Because the time necessary for six tasks would have been too long for one experimental run, a reduced set of tasks was generated with three tasks (one from each domain, which were counterbalanced) each.

The information search process in the task is comparable to a search for information within one of the large search engines like, e.g., Google or Yahoo. The participant reads the task text and enters a keyword into the search field (see Figure 25 for a screenshot of the test environment). After hitting return a list of questions[3] is displayed on the participant's screen which corresponds to the

[3] The fact that the participants' search is actually in a database of questions is the only difference from the above mentioned Google or Yahoo search (standard search paradigm). In such a standard search a

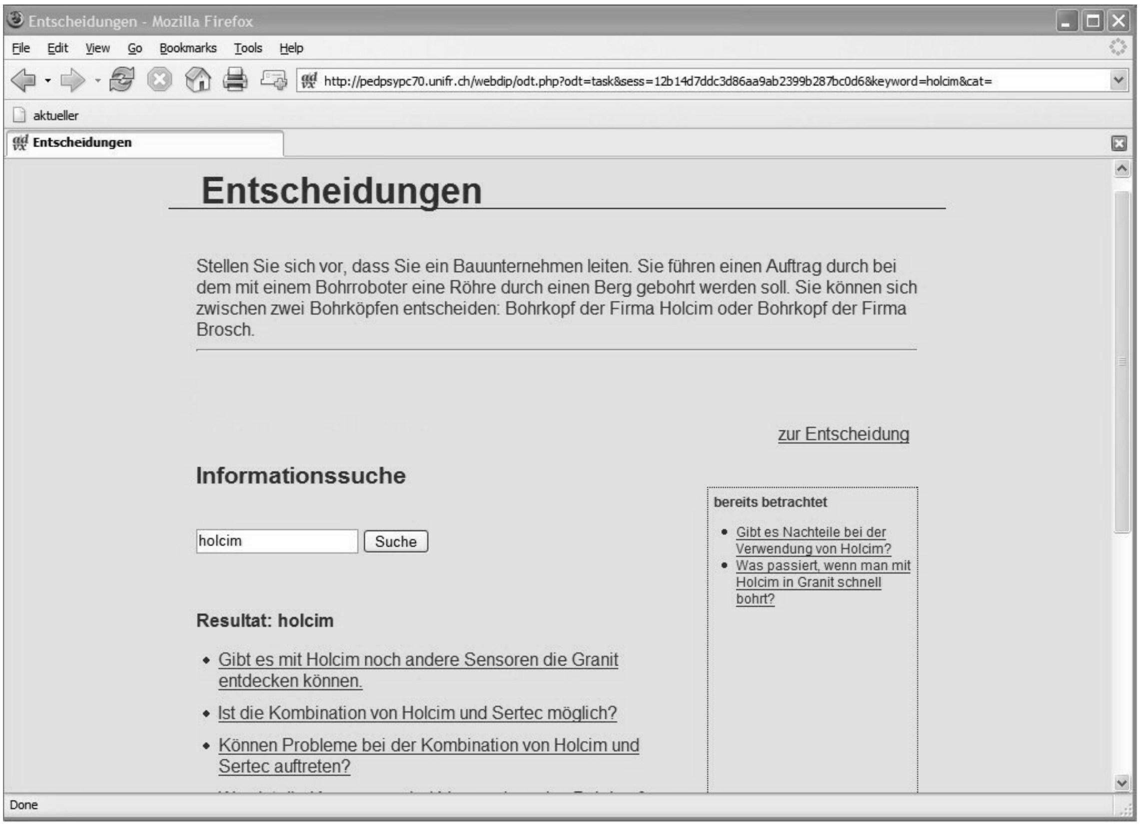


Fig. 25. Participant's view of the keyword search in WebDiP.

entered keyword (see Chapter 7.2.2, p. 69 for detailed information about the handling of keywords within WebDiP). The participant chooses a question of interest, clicks on the corresponding link and receives an answer to his question. Then three possibilities are available to continue with the search: (1) a link for *more information* lists questions with additional information to the one chosen, (2) a *back link* returns the participant to the search results from the previously entered keyword and (3) a *new search* link returns the participant to the former screen to be able to enter a new keyword (without listing of results of the former keyword). On the right hand side of the screen a list with already viewed information provides an "external memory" which can be accessed by clicking on a link within this area for viewing the answer to the question again. The participant can click on the *decide link* at any time within the search process to state the decision for the task. After finishing the experiment the participant is debriefed. At the end of the whole experiment (all three tasks finished) a lottery (which is announced in the invitation e-mail) determines the winners of the following prizes: 2x CHF 150.-, 2x CHF 100.- and 2x CHF 50.-.

---

list of links with short text descriptions of the referenced page is displayed on the results screen after entering of a keyword. This setup can not be used in an information search approach because the only behavioral measure taken from the participants is a click on a link. There are two reasons for this: (1) In a standard search situation it would be hard to determine which information should be presented additionally to the link. (2) Furthermore there would be no way to find out where the participant is actually looking. Using the intermediate step of presenting a list of questions first, solves this problem.

8.2.2 Results

The logic of the data analysis is as follows:

(1) Distribution: An analysis of the distribution of clicks over all tasks and per task is presented, first[4]. Extreme values are identified and eliminated from the data. In reference to their click amount (number of inspected information) groups of participants (low, medium and large number of clicks) are defined for later analysis. A manipulation check concerning the matching of alternatives in the tasks is performed, too (see: *Distribution of clicks and identification of extreme values*, p. 83).

(2) Fusion: To be able to fusion the two tasks belonging to a domain (see Figure 23, p. 79) it has to be shown that tasks yield comparable results concerning amount/type of selected items and/or categories. Therefore the tasks are described and compared from several perspectives, i.e., mean amount of clicks per task, "hits" of clicks per category (see below) and mean amount of clicks per category. The keywords used in the searches are analyzed in this section, too (see: *A mean view of the data*, p. 85).

(3) Sequence: On the item level the tasks will be analyzed concerning different patterns and/or sequences in the participants' clickstream. This step should provide insight into the participants' strategies (see: *Patterns in the clickstream*, p. 93).

Distribution of clicks and identification of extreme values

An analysis of the overall distribution of clicks (without taking into account any manipulation) should reveal insight on the general search behavior between participants. Given that a larger number of clicks can be ascribed to larger interest in the experiment different groups should be distinguished for some of the following analysis.

The overall mean of clicks found was $M= 17.88$ (SD=14.59) with a skewness = 1.40, i.e., a distribution skewed to the left (see Figure 26). The median of the click distribution is at 14, 90% of participants clicked less than 38 times. While there were 22 participants with only 1 information item looked at (click), there were 40 with more than 50. An arbitrary cutoff point is set at the 9th percentile, participants above this point (an overall number of 35 or 8.96%) are excluded from further analysis (through this exclusion the number of participants is reduced to 333). Hawkins (1980, p. 1 as cited in Osborne & Overbay, 2004) defines an outliers as "deviating so much from other observations as to arouse suspicions that it was generated by a different mechanism". The 9th percentile represents such a large deviation from the rest of the sample. This reduction is done for each task and appendant frequency distribution (see a per task histogram in the Appendix on p. 132).

Additionally to the outlier analysis participants are grouped for later analyses, according to their click rate (amount of searched information) per task into a *small*- (within the 3rd percentile) a *medium*- (between the 3rd and the 6th percentile) and a *large*-search group (between the 6th and the 9th percentile). The group building an the beforehand exclusion of participants is a necessary step to account for any behavior that is measured with mean amount of clicks. An outlier in the 9th percentile who had, e.g., 80 clicks would cover a majority of the effects because of overweighing of such a large mean in the corresponding analysis. However, for analysis on the sequence level clearly a disadvantage of this approach is, that a longer search provides a higher chance of finding sequences in the search process.

[4] The necessary data checks and a description of transformation steps from the log-file to a processable file into SPSS is presented in the Appendix A.6, p. 130

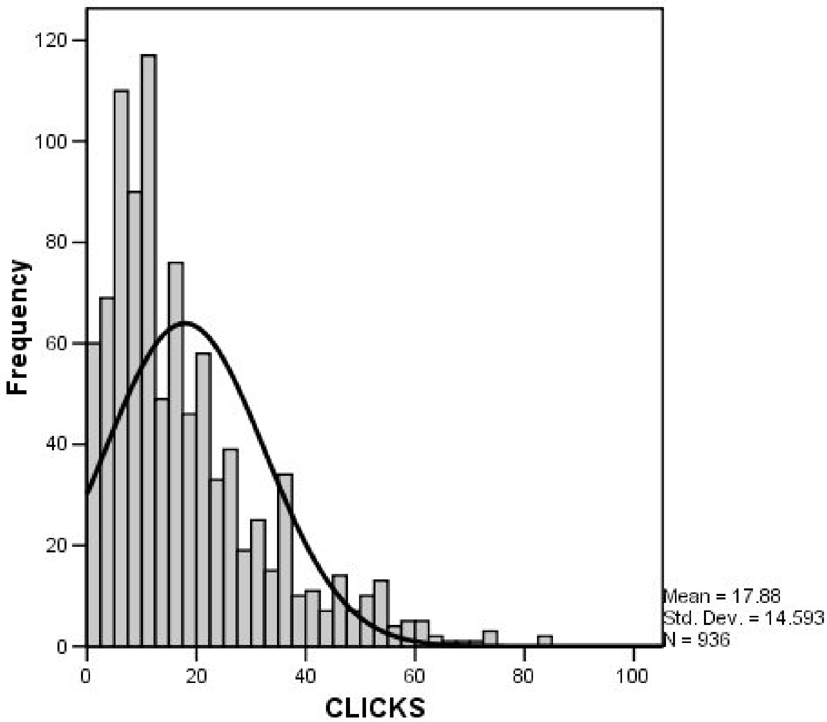


Fig. 26. General distribution of clicks

Manipulation check task position

It seems important to inspect whether the position of a task which was varied with a Latin-square (see above) would have an effect without this control mechanism. A change over positions, i.e., time could be interpreted as learning effects through better understanding of the tasks' structure.

Each participant worked with three of the six available tasks. These three tasks were randomly assigned but controlled for the precondition that every participant received exactly one task from the three domains. A decrease from POS 1 to POS 3 in the amount of inspected information was found with POS 1: M = 15.58 (SD = 10.81), POS 2: M = 14.46 (SD = 9.58) and POS 3: M = 13.67 (SD = 9.66). However this decrease did not reach significance in an ANOVA with amount of click as dependent and POS as independent variable $F(1{,}849)$ = 2.63, n.s.. Nevertheless on a descriptive level it is possible to observer a decrease in clicks (i.e., interest) in the task given a later time in the experiment.

Manipulation check choices

The six tasks were constructed following one schema depicted in Figure 24, p. 80. The two alternatives in this schema (a and b) were matched in order that no alternative outperforms the other.

This was done to keep the search process up as long as possible. To check for the effectiveness of this construction process the analysis of choices between the conditions should yield *no* systematic preference for one alternative.

Table 14. Percentage of choices per task

task	choice	%
drug	Spinox	38.6
	Lofa	61.4
jail	Hinterheim	42.4
	Tierau	57.6
drilling	Brosch	40.6
	Holcim	59.4
operation	Inselspital	49.2
	Südklinik	50.8
rehabilitation	Lenkheim	51.8
	Seewald	48.2
pipe	Bucher	52.8
	Steiner	47.2

Results for this analysis are presented in Table 14. The largest differences between alternatives were found for the drug, jail and drilling task. Here only 38.6%, 42.4% and 40.6% (resp.) of participants chose the first alternative. For the other three tasks (operation, rehabilitation and pipe) choices closely matched for both alternatives. Despite the small preferences for alternative *b* in the first three tasks a domain consideration from this perspective is still possible because the second group of tasks (operation, rehabilitation and pipe) resulted in balanced choices between alternatives.

A mean view of the data

The two main factors of the experiment – CUE (2, 3 or 4 cues) and DOMAIN (Medical, Law, Business) are inspected next. The three CUE levels result in a lower mean click rate for the 2 cues: $M_2 = 13.88$, $SD_2 = 9.55$ than for the 3 cues: $M_3 = 15.07$, $SD_3 = 10.55$ or the 4 cues: $M_4 = 14.82$, $SD_4 = 10.09$.

The domains are generated through a combination of the corresponding tasks: drug and operation (op) → MEDICAL (MED), pipe and drilling (drill) → BUSINESS (BUS), jail and rehabilitation (rehab) → LAW. The mean click rate for the MED domain is the lowest with $M_{MED} = 13.74$, $SD_{MED} = 8.68$. For BUS and LAW larger means are found, which still are close together: $M_{BUS} = 14.48$, $SD_{BUS} = 10.18$ and $M_{LAW} = 15.51$, $SD_{LAW} = 11.12$ (see Table 15, 16 for an overview).

A two factorial ANOVA with amount of clicks as dependent and CUE (2,3,4) X DOMAIN (BUS, MED, LAW) as independent variables was administered. No main effect of CUE and DOMAIN as well as no interaction reached a significant difference (all $F's < 1$).

On the next level the mean clicks per task are used as a dependent measure. This step is taken in order to check for the fulfilment of the precondition that two tasks from one domain should be equal in terms of considered information. The analysis should shed more light on the zero-result above. *Operation* results in the lowest amount of clicks ($M_{op} = 13.04$) whereas *rehabilitation* in the largest ($M_{rehab} = 16.55$). In between this range *drug, pipe, drilling* and *jail* reside between a mean of $M_{pipe} = 13.54$ and $M_{drilling} = 15.35$. The difference of means between tasks (which is relevant

Table 15. Mean Clicks for Domains and Cues

DOMAIN	CUE	Mean	SD
MEDICINE	2 cues	12.39	8.23
	3 cues	15.76	9.66
	4 cues	13.18	7.79
LAW	2 cues	14.66	10.30
	3 cues	15.39	11.44
	4 cues	16.43	11.62
BUSINESS	2 cues	14.63	9.93
	3 cues	14.04	10.54
	4 cues	14.73	10.17

for the grouping of tasks to domains) is smallest for the MED domain with $d_{drug-op} = 1.38$, i.e. the tasks in the MED domain show the largest similarity in terms of mean amount of clicks. In the LAW and BUS domain there is a larger mean difference of $d_{drilling-pipe} = 1.81$ and $d_{rehab-jail} = 2.22$ respectively (see Table 16 for an overview of means per DOMAIN and TASKS).

Table 16. Mean Clicks for Domains and Tasks

Domain	Mean	SD	Task	Mean	SD
MEDICINE	13.74	8.68	operation	13.04	8.49
			drug	14.42	8.83
BUSINESS	14.48	10.18	pipe	13.54	8.92
			drilling	15.35	11.18
LAW	15.51	11.12	jail	14.33	9.73
			rehabilitation	16.55	12.11

An ANOVA for the dependent factor amount of clicks and the independent factor TASKS (drug, jail, drilling, operation, rehabilitation, pipe) results in a significant main effect $F(5,846) = 2.31$, $p < .042$ for TASKS. Despite this effect no significant differences between tasks in a post-hoc Scheffè test could be found. The main effect probably results from the relatively large difference between operation and rehabilitation. However this difference averages out when the other tasks are added. Given that the amount of clicks is an indicator for interest/hardness of a task, these results about the generation of the tasks for the three domains seem to have some problems concerning their allocation.

The above analysis could result because of a skewed distribution of clicks per category (a detailed analysis on the level of categories can be found below!). Considering the means per category for the six tasks (see Table 19, p. 91) especially categories ranging close to *0* like *general probability* or *probability of subsequent action a1* could influence the above results seriously. To account for this influence data were analyzed again using a *hit per category* measure. A hit is generated as soon as *one* (and only one) item in a category is chosen. Because of the relatively large number of participants (and clicks) it is possible to run a two factorial ANOVA again using *hit per category* as dependent and CUE (2,3,4) X DOMAIN (BUS, MED, LAW) as independent variables. Again no main effect or interaction reached a significant difference (all $F's < 1$).

Category level

Apart from the quite superficial level of means across conditions a more detailed impression of the data can be gained from looking at the distribution of clicks for categories of information. Each item was rated by two independent rates ($r = .95$) and categorized into one of 13 categories (see Table 17).

Table 17. The 13 categories of information

No.	category name	short description
1	general information	general information about alternative a and b
2	P_a	probability information about alternative a
3	P_b	probability information about alternative b
4	alt *a*	information about alternative a
5	alt *b*	information about alternative b
6	rdo	other risk defusing operators
7	cons	consequences of an alternative
8	new alt	new alternatives additionally to a and b
9	general P	information about basic probabilities
10	P_{a1}	probability information about subsequent action a1
11	P_{b1}	probability information about subsequent action b1
12	*a1*	information about subsequent action a1
13	*b1*	information about subsequent action b1

In order to get a clearer picture about the categories an example for each of the categories presented in Table 17 in short form will be given now. The operation task is chosen for these examples because the content of this task (the choice between two operations in two different hospitals) is easy to understand. Category 1 – *general information* – includes items that describe properties of both alternatives like: “Which hospital is better?” or “What are the symptoms of the illness?”. Categories 2/3 – *Probability of alternative a/b* – provides probabilities of events resulting from alternative a/b: “What is the probability that operation a will be a success?” or “What is the probability that operation b is a failure?”. Categories 4/5 – *information about alternative a/b* – includes all but probability information about the two alternatives a/b: “Are there side effects of alternative a?” or “How is the procedure performed?”. Category 6 – *risk defusing operators* – includes RDOs not included in categories 12/13 (see below): “What screening is additionally available to alternative a/b?”. Category 7 – *consequences* – presents consequence if an alternative is chosen: “What consequences does the operation in hospital a have?” or “Are there any other side effects of alternative b?”. Category 8 – *new alternatives* – describes new alternatives that could be done instead of alternative a/b: “Are there other possibilities to stop the bleeding?”. Category 9 – *general probability* – provides probabilities not directly connected to alternative a/b or a1/b1: “What is the prevalence of a tumor in the stomach?”. Categories 10/11 – probability of subsequent action a1/b1 – specific probabilities about the subsequent actions a1 and b1: “What is the likelihood that an endoscopy can find the bleeding (subsequent action a1)?” or “What is the likelihood that the test shows a bleeding even if there is non (a false positive)?”. Categories 12/13 – *information about the subsequent action a1/b1*: includes all but probability information about the two subsequent action a1/b1: “What can I do against the side effects of the operation a1?” or “What is an endoscopy?”.

Taking a look at the mean clicks per category Figure 27 summarizes these for the 13 categories. A one factorial ANOVA with mean amount of clicks as dependent and CAT (general info, P alt a,

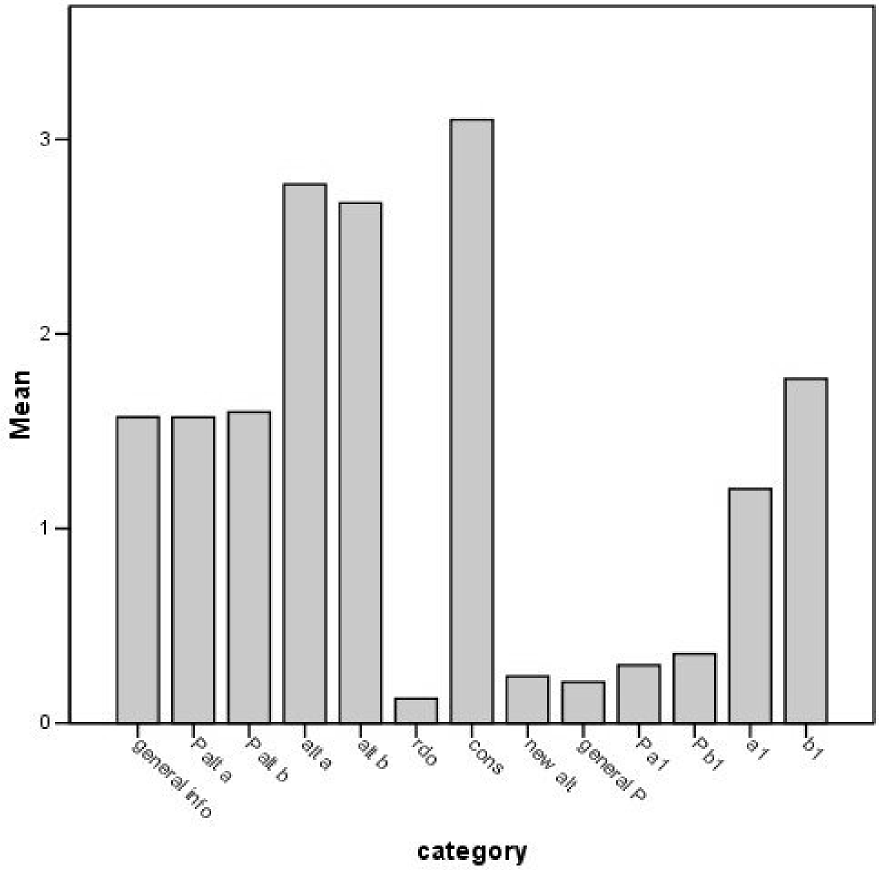


Fig. 27. Mean clicks per category for all tasks

P alt b, alt a, alt b, rdo, cons, new alt, general P, P a1, P b1, a1, b1) as independent variable results in a significant main effect $F(12{,}12130) = 305.97$, $p < .000$. Three categories have a very frequent usage – these are alt *a*/alt *b* and *cons* (all three reach significantly higher values in comparison to the other 10 categories, after Bonferroni adjustment for multiple comparisons with $p < .001$). The usage of information about the categories *a* and *b* seems natural due to the fact that these are among the last information items participants receive in the tasks texts. The frequent usage of these categories can additionally be seen when search terms are inspected in the next section in Table 20. Also on the search term level the two alternatives clearly get the highest attention (see a detailed analysis for search terms below).

After these three front-runners in interest the following categories reside in the midfield of interest: general information, P alt *a/b*, *a1/b1* (no significant differences were found between these categories). The probability information for alternative *a* and *b* is of moderate interest to the participants. This pattern of – large interest in alt *a/b* and moderate interest in P alt *a/b* – can be found again for *a1* and *b1* where only moderate interest is shown in the alternatives and even small interest in the connected probabilities (P *a1/b1*). Only small interest was also shown for rdo, new alternative, general P and P *a1/b1* (no significant differences are found between these categories). The rdo category is under represented in this illustration because of the categories *a1/b1* and P *a1/b1* which are (all four) rdos, too. However *new alternatives* and *general P* are of an overall small interest for the participants of this study.

Table 18. Test statistics for the 12 significant categories

No.	CAT	df	df_{error}	F-value	$p <$ ...
1	general information	5	928	6.52	.00
2	P_a	5	928	11.97	.00
3	P_b	5	928	6.13	.00
4	alt a	5	928	4.97	.00
5	alt b	5	928	4.11	.00
6	rdo	5	928	5.10	.00
7	cons	5	928	12.98	.00
8	new alt	5	928	5.34	.00
10	P_{a1}	5	928	6.79	.00
11	P_{b1}	5	928	2.55	.026
12	*a1*	5	928	2.78	.017
13	*b1*	5	928	5.21	.00

A second analysis wants to gain more insight on the differences between tasks on the category level. The question for this analysis is whether the categories result in different levels of differentiation between tasks. For each of the 13 categories (CAT) an ANOVA was calculated. The independent measures[5] were the mean clicks (for separated categories, i.e., 13 different calculations) the dependent measure the TASK (drug, operation, jail, rehabilitation, drilling, pipe). Every CAT with the exception of *general probability (9)* gained a significant main effect between tasks (the test statistics of the 12 remaining categories are reported in Table 18).

The post-hoc tests of the significant main effects are summarized beginning with the largest differences of CAT over TASKs and moving on the smaller, less important differences.

The following CATs revealed the largest differences of mean amount of clicks between tasks: *general probability, probability of alternative a, consequences and b1* (see Figure 28 for line charts of these four categories) and are therefore closer inspected.

Consequences show mean differences between 1.50 (rehab – pipe) and 1.9 (drug – op) ($p < .05$). These are the largest differences found for all CATs. Clearly two groups can be separated within this CAT, group 1: op and pipe are low in interest – the consequences for these two tasks are clear to the participants. Group 2: drug, drilling, rehab are considerably higher – consequences of the alternatives within these tasks are less clear to the participants. In between those two groups jail also receives relatively high mean click rates – pointing again at a larger interest in the consequences of this task. The picture for the other three CATs is not that accentuated. *General info* results in significant differences between drug – drilling/rehab as well as drilling – op. All of them are between 1.12 and 1.26 clicks ($p < .05$). Drilling and rehab show a larger need for general information than the other tasks. In the two MED tasks (drug, operation) hardly any general information is needed at all.

Probability of alternative a results in significant difference between the pipe and the five other tasks. In the pipe task significantly less information about *probability of alternative a* (i.e., the company which repairs without additional tests) is of interest than in the other tasks. The pipe task is extremely low within probability of alternative a – probability information seems to have no influence on choosing between two companies.

[5] because of the large number of data from 13 ANOVAs the results will be summarized

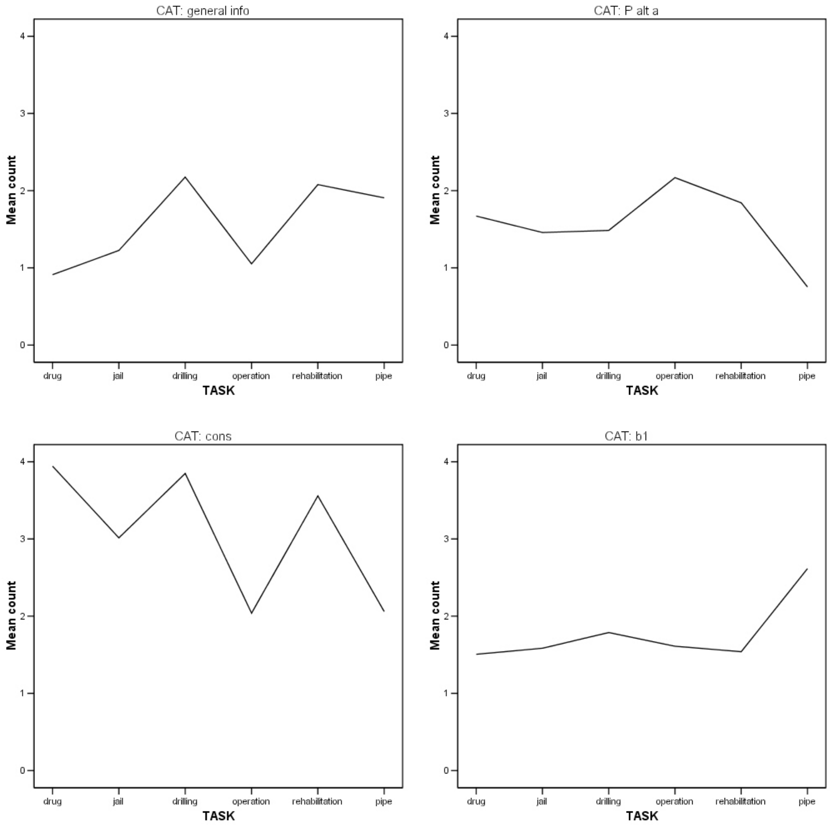


Fig. 28. Means of pairwise comparisons between categories and tasks

The fourth CAT – *b1* – includes information about the additional action to alternative *b*. With *b1* the pipe task differentiated best between tasks with significant more information requested than the drug, jail, op or rehab task. The significant pairwise comparisons are indicated by superscripts $^{a,b,c...}$. Inspecting *a1* and *b1* in Table 19 shows that the pipe task results in an outlier for b1 with more than double the amount of searched for information than with the other tasks. This alternative is much more interesting in the pipe task than in all the others. For an overview of the mean clicks for the 12 categories[6] and six tasks see Table 19.

Adding the CUEs to categories

The main question for the next analysis is whether the two tasks assigned to one domain produce comparable patterns in reference to the categories. This question is analyzed in terms of a comparison of all 13 alternatives for the three introduced CUE levels (low, medium and high amount of information). In order to get a better overview of category usage between tasks, clicks for the six tasks and corresponding categories are plotted in the Appendix in Figure 37, 38 and 39, p. 134 - 136. The impression delivered by the overview above (Table 19) can be found in these plots again. In the LAW domain (the middle two plots) relatively homogenous search behaviors can be

[6] category 9: *general P* did not result in significant differences

Table 19. Mean clicks for categories per task

No.	category	drug	jail	drill	op	rehab	pipe
1	general info	$0.91^{a,b}$	1.23	$2.18^{a,c}$	1.05^{c}	2.08^{b}	1.91
2	P alt a	1.67^{a}	$1.46^{b,c}$	$1.48^{d,f}$	$2.17^{b,d,e}$	1.84	$0.75^{c,e,f}$
3	P alt b	1.70	1.37^{a}	1.44^{c}	$2.07^{a,b,c}$	1.82^{d}	$1.15^{b,d}$
4	alt a	3.01	2.63	2.80	2.21^{a}	$3.39^{a,b}$	2.51^{b}
5	alt b	2.75	2.77	2.64	2.05^{a}	3.16^{a}	2.64
6	rdo	0.16	0.08^{a}	$0.24^{a,b,c}$	0.05^{b}	0.13	0.07^{c}
7	cons	$3.94^{a,b}$	3.01	$3.85^{c,f}$	$2.04^{a,c,d}$	$3.56^{d,e}$	$2.06^{b,e,f}$
8	new alt	0.07^{a}	0.28	0.26	0.19	0.24	0.40^{a}
10	P a1	0.16^{a}	0.18^{b}	0.33	0.30^{c}	0.24^{d}	$0.57^{a,b,c,d}$
11	P b1	0.18	0.43	0.48	0.15	0.52	0.37
12	a1	1.21	1.15	1.21	1.56	1.05	1.03
13	b1	1.51^{a}	1.58^{b}	1.79	1.61^{c}	1.54^{d}	$2.62^{a,b,c,d}$

The superscripts $^{a,b,c...}$ denote that on the corresponding category a significant difference ($p < .05$) was found in a pair-wise comparison between tasks with the same letter. For example in the general info category (1) tasks drug-drill(a), drug-rehab(b) and drill-op(c) reached significant differences.

observed. In the other two domains clearly a shift in clicks for drug and drilling in comparison to operation and pipe are found (compare to the mean values for tasks in Table 16, p. 86).

From a statistical point of view Fisseni (2004) suggests that for the comparison of profiles (the distance between profiles) a *product moment correlation* (r) should be calculated. Given the hypothesis that corresponding tasks should have (more or less) equal profiles (i.e. equal means between categories) the correlations for these tasks – drug/operation, jail/rehabilitation and drilling/pipe – should be higher than the correlation of these profiles with the other tasks. The bar charts for TASK X CAT are plotted in the Appendix p. 134 – 136. An inspection of the correlation matrices in Figure 43 reinforces the impression one gets when the bar charts are considered visually.

The correlations for e.g., drug/operation as well as jail/rehabilitation are high (e.g., $r_{drug-op}$ = .70 (2 cues condition), $r_{jail-rehab}$ = .87 (3 cues condition) respectively). The correlations between the target and the other tasks are high (or even higher), too (e.g., $r_{pipe-drug}$ = .76 (4 cues condition), $r_{jail-drill}$ = .88 (3 cues condition)). Therefore it is not possible to differentiate two task from a suspected domain from the other tasks. In order to account for the different absolute heights of the profiles (the differences in means of clicks per task) the *intra class correlations* were calculated, too. A comparison of this analysis to the current product moment correlations did not reveal sufficient differences (again tasks from one domain showed high correlations but also between domains (e.g., pipe and drug) correlations were high(er)).

Because of the large number of categories (13) it was suspected that underlying patterns could be hidden through the detailed clustering of the data. Therefore the comparison of profiles was done again for a reduced set of categories. The following categories (CAT_{red}) were chosen: general information/consequences (these categories showed relatively large variations in the first analysis) as well as a the four central categories from the task structure (see Figure 24, p. 80) P_{altab} (Probability of alternative *a/b*), alt_{ab} (alternative *a/b*), P_{a1b1} (Probability of subsequent action *a1/b1*), alt_{a1b1} (subsequent action *a1/b1*). The profiles for TASK X CAT_{red} are plotted in the Appendix p. 137 - 139. A visual inspection of the profiles leads to the assumption that in all three cue conditions the middle two profiles – jail and rehabilitation (LAW domain) – match each other nicely. For the other two domains such a match can not be found. An inspection of the correlation matrices for the three CUE conditions (see Figure 44) strengthen this impression. The correlation $r_{jail-rehab}$ = .99 is high (and even highest in this domain) but unfortunately this is only true for the 2 cues condition. In the 3 cues condition $r_{jail-rehab} = .89$ is still high but $r_{jail-drug} = .99$ reaches even higher coverage. This pattern (high correlations between two tasks from one domain but also high, or even higher, correlations for tasks from one domain with another) can be found in the data over and over again. The additionally performed *intra class correlation* did not reveal new insights in this case, too.

Table 20. Percentage of entered keywords per task

Task	comp[a]	keyword	%	Task	comp	keyword	%
DRUG	*b*	lofa	30.0	OP	*b*	südklinik	27.9
	a	spinox	28.4		*a*	inselspital	26.0
		nebenwirkungen	9.1			operation	6.9
		medikament	3.3			blutung	2.2
	b1	rigolin	3.1			klinik	2.1
	a1	tremol	2.9		*a1*	endoskopie	2.1
		kosten	1.7			gewächs	1.7
		kopfschmerzen	1.5			kosten	1.7
		wirkung	1.5			spital	1.7
		gehirnkrankheit	1.4		*b1*	magenspiegelung	1.7
		berioenzephalitis	1.1			arzt	1.3
		krankheit	1.0				
JAIL	*b*	tierau	34.0	REHAB	*b*	seewald	32.9
	a	hinterheim	33.8		*a*	lenkheim	31.4
		gefängnis	8.5			gefängnis	6.8
		ausbruch	6.4		*a1*	sozialprogramm	3.3
		sicherheit	3.3		*b1*	gemeinnützig	2.8
	a1	einzelhaft	1.5			unterschied	2.4
		kosten	1.4			zusatzmassnahme	2.1
		wahrscheinlichkeit	1.2			rückfall	1.8
	test	bewährungstest	1.2			kosten	1.6
						straftat	1.5
						urteil	1.2
DRILL	*a*	brosch	25.2	PIPE	*a*	bucher	30.3
	b	holcim	21.5		*b*	steiner	29.1
		bohrkopf	10.9			kosten	7.2
		granit	4.9			wasserrohrbruch	7.1
	b1	sertec	4.8			firma	5.1
	a1	varim	4.5		*a1*	komplettsanierung	2.3
		kosten	3.7		*b1*	totalsanierung	2.1
	test	sensor	2.1			zeit	2.1
		berg	1.5		*test*	magnetometer	1.5
		gestein	1.0				

[a] comp = component of general task structure

Search terms

The search terms entered by the participants were analyzed according to their proportion of nomination of all keywords for a task. Because the spread of keywords is quite large for the tasks only those keywords are reported which exceed 1% of usage. The coverage of keywords using this criterium is around 80%, the remaining 20% are keywords only entered between one and fifteen times (given that the 333 participants used $M_{keywords} = 44.21$, $SD_{keywords} = 8.23$ on average).

Table 20 depicts the percentages of keywords entered per task. Given the general structure of the tasks (see Figure 24, p. 80) the usage of keywords directly relates to this structure. In all but the drilling task (with 46.7%) more than 50% of the keywords were related to alternative *a* or *b* (Spinox/Lofa, Inselspital/Südklinik ...). Between these two alternatives *a* received a higher interest than *b* in drilling and pipe (the BUSINESS domain). In the four other tasks *b* was of highest interest. The second component for each alternative (*a1*, *b1*) was searched for in all tasks but jail (here only *a1* was of interest with 1.5%). The range for the five remaining tasks for *a1* and *b1* was between 1.7% and 4.8% with fairly equal distributions between the two. Remarkable is the fact that in all tasks there was an interest in costs (kosten), this interest was highest in the BUSINESS domain (drilling and pipe, 3.7% and 7.2% resp.).

Table 21. Total number of keywords

	drug	op	jail	rehab	drill	pipe	Total
2 cues	892	692	569	940	968	684	4745
3 cues	802	738	821	888	1071	589	4909
4 cues	748	1023	689	1075	837	897	5269
Total	2442	2453	2079	2903	2876	2170	14923

The total number of keywords entered per task is illustrated in Table 21. Rehabilitation and drilling resulted in the largest usage of keywords (2903 and 2876 respectively). Furthermore drug and op (2442 and 2453 resp.) as well as jail and pipe (2170 and 2079 resp.) closely match in amount of entered keywords.

Patterns in the clickstream

The above described approaches of aggregating process data to identify differences has certain limitations. Although differences between categories or keywords can be demonstrated the interpretation of these data does not allow descriptions about actual strategies in the information search. For the question: "what are *frequent pathes in* or *typical patterns* through the given information?" other approaches are necessary.

Three steps are introduced now that should shed light on the above questions. At first simple transactions between categories of information are described. Then the most frequent transactions are identified and as a final step, on the "lowest" level, the first 10 clicks and the mean ranks of categories are inspected.

A first look at sequences - transactions

A transaction between categories is defined as the switching from one category (e.g., probability[7]) to another (e.g., consequences). This step is indicated by a click on an item from the category *P* first followed by a click on an item from the category *cons.* Tables (22, 23, 24) list the transactions for all 13 categories and 6 tasks (two tasks, representing one domain are listed together).

The following example should ease readability of the subsequent analysis. In table 22 the left most "78" indicates that there were 78 instances where at first an information from category 1 (i.e., *general information*) was selected and the following category was again "*general information*". This will be called a *within category transaction (*wct*)*. In the next column the "3" indicates three

[7] see Table 17, p. 87 for an overview of categories

transactions from category 1 to category 2 (i.e., *general information* TO *probability information about alternative a*) were done. This will be called a *between category transaction (*bct*)*.

Table 22. Transactions Medicine

drug	1	2	3	4	5	6	7	8	9	10	11	12	13	op	1	2	3	4	5	6	7	8	9	10	11	12	13
1	78	3	5	0	3	7	11	2	0	0	0	1	6	1	67	9	4	13	10	2	9	7	0	2	0	11	1
2	1	115	16	60	6	0	17	1	6	0	1	16	0	2	1	160	45	69	14	0	12	0	0	1	1	13	1
3	0	33	114	12	58	0	14	0	3	1	0	2	7	3	5	23	148	7	68	0	15	1	5	2	0	1	7
4	2	53	6	163	39	3	97	1		1		63	8	4	3	35	3	96	77	0	20	1	0	3	1	63	3
5	1	3	40	71	115	1	66	0	0	0	0	3	72	5	11	9	31	41	76	1	17	3	0	0	0	7	44
6	8	0	0	1	0	1	11	0	0	0	0	0	0	6	1	0	0	0	0	1	2	0	0	0	0	1	0
7	20	30	45	81	75	7	250	0	2	0	0	18	32	7	17	46	38	30	12	1	103	0	0	4	2	16	17
8	5	0	0	0	0	0	3	1	0	0	0	0	0	8	7	0	0	3	0	0	1	0	0	0	0	2	0
9	3	2	1	0	0	0	1	1	12		1	1	0	9	1	0	0	3	0	1	2	0	5	0	0	0	0
10	0	0	0	0	2	0	0	0	0	9	1	13	0	10	0	1	0	2	0	0	2	0	0	19	0	15	6
11	2	1	1	1	1	0	0	0	1	0	19	0	0	11	0	0	1	0	2	0	0	0	1	0	15	1	2
12	3	6	1	47	6	0	34	0	0	12	1	48	8	12	3	21	3	17	3	0	54	1	0	11	0	76	33
13	5	1	12	7	40	0	40	0	0	2	4	16	87	13	7	2	15	7	15	0	46	2	0	4	3	13	94

Table 22 lists the transactions for the medical domain with the tasks drug and operation. The most frequent transaction in the drug task is a *wct* in *cons* (category 7, clicks 250)[8], followed by *wct alt a* (4, 163) and the P_{ab} as well as *alt b* (2, 115; 3, 114; 4, 115). For the operation task the *wct* for P_a (1, 160) is the most prominent one, followed by *cons* (7, 103) and *alt a* (4, 96).

The *bct* were most prominent (in the drug task between *alt a* and *cons* (4-7, 97)[9] as well as *alt b* and *b1* (5-13, 72). In the operation task frequent transactions were between P_a and *alt a* (2-4, 69) as well as *alt a* and *a1* (4-12, 63).

Table 23. Transactions Business

drill	1	2	3	4	5	6	7	8	9	10	11	12	13	pipe	1	2	3	4	5	6	7	8	9	10	11	12	13
1	184	6	2	15	6	11	50	23	11	0	8	7	9	1	38	6	7	47	43	4	39	9	9	9	2	6	27
2	5	100	17	57	9	2	9	1	2	0	4	7	1	2	3	12	47	4	17	0	2	0	0	10	1	5	5
3	2	25	101	8	43	0	13	2	0	0	5	1	11	3	8	35	20	11	9	0	3	1	17	10	21	4	16
4	3	59	3	156	80	2	75	0	1	2	0	32	4	4	45	11	8	97	62	0	50	4	2	4	0	18	35
5	9	3	60	45	141	1	63	1	0	1	1	8	46	5	35	6	13	64	87	0	40	24	2	7	2	13	72
6	10	0	0	3	1	0	12	0	1	2	0	3	5	6	1	0	0	3	0	0	1	1	0	0	0	1	2
7	53	11	22	63	42	16	270	4	0	2	0	50	45	7	56	10	7	28	34	1	75	5	1	8	1	23	36
8	20	0	0	0	0	3	6	2	0	1	0	1	5	8	8	1	1	6	6	0	4	1	0	1	1	8	19
9	9	1	1	0	0	0	3	0	12	0	1	1	1	9	5	6	11	2	1	0	1	1	4	2	0	1	6
10	0	0	0	2	2	1	4	1	0	20	0	11	10	10	5	3	1	4	4	0	8	1	0	11	0	7	36
11	11	1	1	1	0	2	1	2	2	0	49	1	1	11	1	3	20	4	2	1	2	0	2	2	13	2	2
12	8	16	0	44	4	1	21	0	0	16	1	43	23	12	12	6	10	22	12	0	11	2	3	3	5	15	42
13	6	1	14	11	39	1	30	2	0	10	6	24	116	13	46	7	19	29	81	0	25	6	2	14	9	45	73

Table 23 lists the transactions for the business domain with the tasks drilling and pipe. The most frequent transaction in the drilling task is a *wct* in *cons* (7, 270), followed by *wct general information* (1, 184), P_{ab}, *alt ab* and *b1* (2, 100; 3, 101; 4, 156; 5, 141; 13, 116). In the pipe task only *alt a* was chosen relatively often (4, 97).

The *bct* in the drilling task were between *alt a* and *cons* (4-7; 75) as well as *alt b* and *cons* (5-7; 63). Less pronounced was found for the pipe task with transactions between *cons* and *general information* (7-1; 56).

Table 24 depicts the transactions for the law domain with the tasks jail and rehab. The most frequent transaction in the jail task are two *wct* in *alt b* (5, 152) and *cons* (7, 148), followed by *wct alt a* (4, 122). In the rehab task *cons* was chosen most often (7, 233). Of large interest were also the categories *general information*, P_{ab} and *alt ab* (1, 124; 2, 156; 3, 155; 4, 193; 5, 155).

[8] the first digit (*wct*) in brackets refers to the corresponding category and the second digit refers to clicks per transaction

[9] the first two digits (*bct*) in brackets refer to the two categories and the second digit refers to clicks per transaction

Table 24. Transactions Law

jail	1	2	3	4	5	6	7	8	9	10	11	12	13	rehab	1	2	3	4	5	6	7	8	9	10	11	12	13
1	81	0	1	8	3	3	18	12	7	0	6	9	8	1	124	4	5	31	23	3	61	14	14	1	8	12	8
2	0	94	21	44	2	0	13	1	1	1	0	7	1	2	3	156	22	51	16	0	17	1	0	0	0	7	2
3	0	17	87	12	44	0	10	0	0	0	4	0	7	3	9	24	155	10	56	0	16	0	0	1	0	3	7
4	1	36	3	122	47	0	81	0	2	0	0	31	3	4	33	53	3	193	103	2	79	1	0	0	1	35	7
5	6	2	31	35	152	0	73	3	0	0	0	1	27	5	33	6	51	89	155	2	75	9	0	1	0	6	29
6	5	0	0	3	0	0	0	0	0	0	1	0	1	6	2	1	0	1	0	0	2	0	1	8	1	0	4
7	21	29	31	41	36	2	148	3	1	0	0	33	49	7	34	37	38	50	46	3	233	1	7	1	1	39	42
8	18	0	0	3	0	1	0	4	0	0	1	8	4	8	16	0	0	1	2	1	0	5	0	0	0	8	2
9	8	2	1	0	0	0	0	1	10	0	1	0	1	9	14	3	1	0	0	1	1	0	16	0	2	0	0
10	0	0	0	0	1	0	4	0	0	8	1	10	1	10	0	0	0	2	2	0	0	0	0	13	0	1	14
11	0	5	2	2	0	0	1	1	3	1	39	1	2	11	4	3	0	1	4	1	5	3	0	1	53	1	6
12	2	10	0	40	3	2	15	1	1	9	0	37	25	12	9	6	1	44	6	0	19	0	0	6	0	41	26
13	4	9	2	7	10	38	3	27	5	1	6	7	14	13	9	0	5	22	47	2	30	1	0	3	9	11	85

The *bct* in the jail task did not receive high click rates, *cons* and *b1* (7-13, 49) was the most prominent one. The same is true for the rehab task with one prominent transaction between *general information* and *cons* (1-7, 61)

Summing up the results for the six tasks (three domains) a strong tendency for staying within a task category can be observed. As soon as participants found an information type that was interesting to them they stayed with this category. In three out of the six tasks (drug, drilling and rehabilitation) this category is *cons*. The other three most interesting categories are: P_a, *alt a* and *alt b*. But there is not only a strong trend to stay within the category *cons* also the move *to* this category is performed relatively often. This is especially true for the two medical tasks. The moves from category *cons* concentrate mainly on the central categories P_{ab} and *alt ab*.

Table 25. Initial categories

task/category	1	2	3	4	5	6	7	8	9	10	11	12	13
drug	2	6	13	19	51	5	46	1	1	0	1	2	2
operation	19	15	13	44	29	1	10	4	0	0	0	6	1
jail	20	3	5	39	36	0	14	8	1	0	1	4	7
rehab	33	0	8	45	29	6	22	2	1	0	5	3	1
drilling	31	6	4	36	26	0	51	1	0	0	0	1	0
pipe	12	2	0	40	28	5	36	4	1	3	0	4	8
sum	117	32	43	233	199	17	179	20	4	3	7	20	19

The initial category (the category first selected) is not included in the above description and will be inspected in more detail now. Table 25 lists the first click for the 13 categories and the six tasks. The most prominent starting categories are *alt a* and *alt b*. This was already reflected in table 20 where the main entered keywords were listed - the two categories *alt ab* (4, 5) received the highest percentages there, too. Two further categories were likely requested as initial information - *cons* (7) and *general information* (1).

A second look at sequences - most frequent patterns

This section introduced a *cumulative* approach which shows the *most frequent* patterns in the searches. In the next step, even more detailed, *actual sequences* in the information search will be identified.

The identification of information sequences in databases is a common research problem in informatics. Especially in consumer research the analysis of, e.g., shopping baskets profits from tools that can identify corresponding items in an online shopping process. Two recent approaches to this problem are introduced with the algorithms MAFIA (Maximal Frequent Itemset Algorithm: Burdick, Calimlim, & Gehrke, 2001) and SPAM (Sequential PAttern Mining: Ayres, Flannick, Gehrke, & Yiu, 2002). Both algorithms are Open Source and well documented in an online project which can be accessed at: http://himalaya-tools.sourceforge.net/ (retrieved August 20th, 2004).

MAFIA is an algorithm for mining *maximal frequent itemsets*, i.e., the actual sequence of two items in an itemset is not of interest for this method. An important concept for understanding the introduced algorithms is the *support* an itemset has in a database. *Relative support* is defined as the percentage of itemsets (the occurrence of an itemset) in a database (this measure was used to classify sequences for the following tables), whereas *absolute support* denotes the number of occurrences of a sequence (further on only relative support will be used). Because MAFIA searches for itemsets that occur for more than one participant it only gives us insight on collectively used *target-items* (independent of the number and type of other items between target items).

More interesting in terms of search strategies is the SPAM algorithm as described in Ayres et al. (2002). Here not only the occurrence of items but also their sequence is of interest. A *sequence* is defined as an *ordered* list of itemsets with the size m, the number of itemsets in the sequence. The length of a sequence is defined as:

$$l = \sum_{i=1}^{m} |s_i| \tag{4}$$

The search space for a counting algorithm grows very fast with only few items to choose from. The growth is demonstrated in Figure 29 for a sequence tree of two items a,b with a maximal sequences length of $s_{max} = 3$. For the current experiment not only two but 99 items were used with an $s_{max} = 38$ (limited by the data reduction described on p. 83).

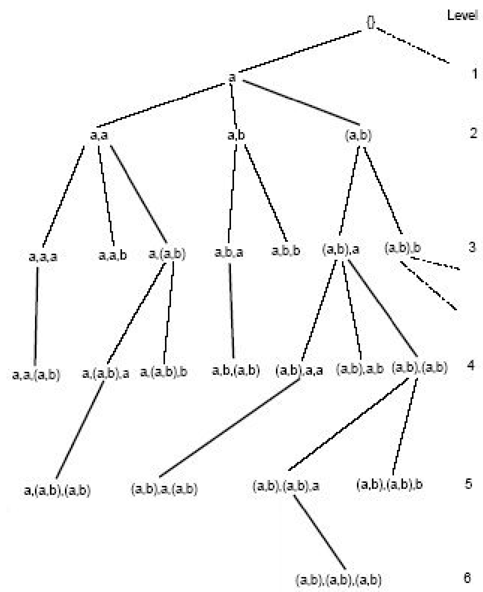


Fig. 29. Sequence tree for two items with $s_{max} = 3$

Several reductions of this search space (e.g., item-set or sequence set pruning - see Ayres et al., 2002 for technical details) are included into the SPAM algorithm that enable an application of the algorithm to large databases without a need for longer processing time.

Table 38, 39 and 40 in the Appendix depict the sequences that reached a support of more than 20%. For this support l has a range of $1 \leq l \leq 5$ in the current experiment. The maximum l is reached in the OP task with 5 whereas all the other tasks only reach $l_{max} = 2$. An interesting measure for how clear the structure of a task is, is the number of frequent sequences with the chosen support value. Three tasks (jail, drill and pipe) generate 38, 41 and 38 sequences respectively. A second group is drug and rehabilitation with 50 and 58 respectively. The outlier task operation generated more than 111 frequent sequences (the support value was set higher for this task because an extremely large number of sequences was found). If these numbers are high (e.g., in the operation task) participants search the database in a comparable way. Contrary if they are low (e.g., in the jail or pipe task) only few commonly used sequences can be found. Given this assumption three groups can be differentiated with *low* (jail, drill and pipe) *medium* (drug and rehabilitation) and *high* (operation) agreement on task structure. Because the above introduced tables do not provide easy insight into actual strategies a presentation method was developed to graphically show which patterns occur (with corresponding percentages). On the x-axis of the Figures 30, 31 and 32 the 13 categories, on the y-axis the support (percentage) for the information items are depicted. Higher support (percentage) means that an information item is selected by a larger number of participants in the current sample.

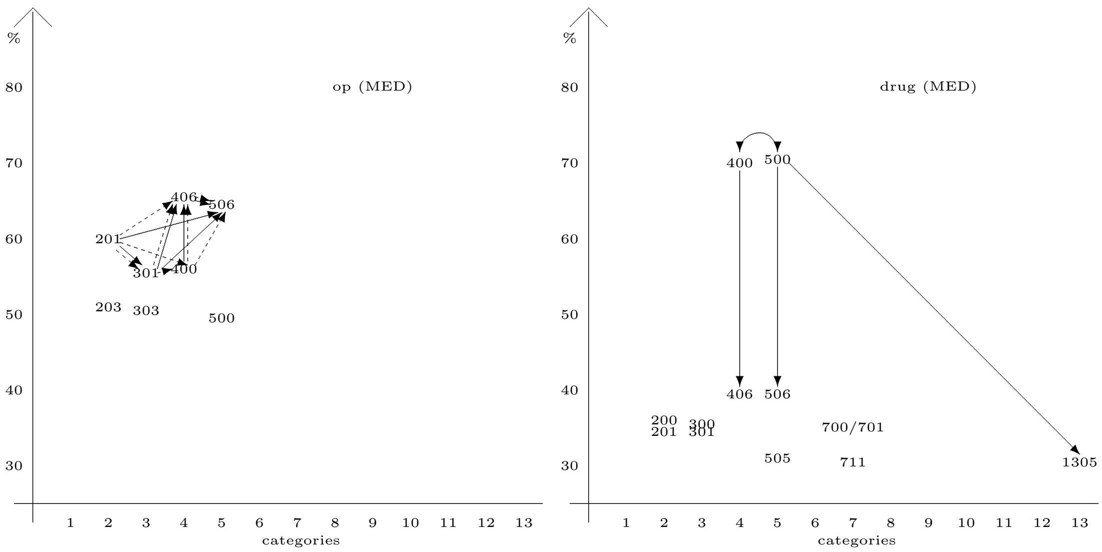


Fig. 30. Patterns for the operation and drug task (MED). Arrows indicate a sequence from item $x \rightarrow y$. Dotted arrows $x \dashrightarrow y \dashrightarrow z$ indicate a sequence longer than $s > 2$

On the a more detailed level the description of the sequences will include the 10 sequences with the highest appearance and selected longer sequences ($l > 1$) for demonstrating strategies in information search. In the drug task (see Figure 30) the side effects of the two alternatives (400 and 500) received the highest ranks (see Table 35 - 36, p. 128 - 129 in the Appendix for the IDs used in Figure 30 - 32 and a list of corresponding questions from the experiment). Following these two (side effects of a and b) the sequence of side effects (a - b i.e., 500 400) was of interest. A "costs" block can be observed next with 406 and 506 (asking for costs of alternative a and b). Still in the upper 30% range are two questions about the "chances of healing" (200, 201, 300, 301) for

a and b; the "consequences of alternative a or b" (700/701); the "side effects" of the combination $a + b$ and finally the side effect of b.

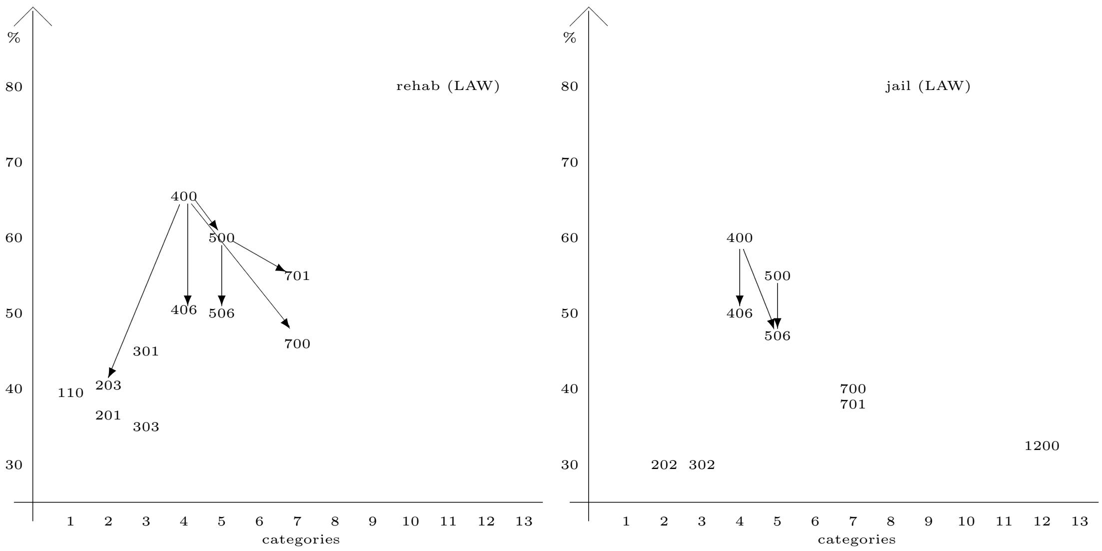


Fig. 31. Patterns for the rehab and jail task (LAW). Arrows indicate a sequence from item $x \rightarrow y$

In comparison to the drug task the operation task (see Table 30) had a lot more items with high support. As already mentioned above this pattern is accredited to the better understanding of the structural demands of the task. Participants agreed more on items which are informative for answering the op task. These items were: costs of alternative a and b (406, 506), the combination of the two and "chances of healing" with alternative a (201) are all above a 60% support. In the 50% block the "side effects" of alternative a (400), again the chances of healing (201, 301), the indications for a "negative event" (203, 303) and side effects of alternative b are listed. The operation task is the only one which resulted in large sequences ($3 \leq s_{drug} \leq 5$). The following are selected for nomination: (1) "chances of healing" of alternative a and b followed by side effects of a and costs of a and b (201 301 400 406 506) - the longest sequences for all tasks. (2) There are several combinations of the items found in sequence (1) like 201 406 506, 201 301 406 506 or 400 506 - the central interest of participants seems to circle around those items.

The two tasks from the LAW domain (see Figure 31) showed comparable interest in information like observed in the above tasks. This interest was concentrated on the first six items – these were (in both tasks): disadvantages of a or b (400/500), costs of a or b (406/506) and consequences of a or b (700/701).

In the jail task unique items are the questions for probability of a negative event in a or b (202/203) and the sequence disadvantages and cost of a (400 406). Contrary in the rehabilitation task the probability of "everything is in order" with option b (301) and "the possibility of combining" a and b (110) were used with more than 30% support.

For the last two tasks from the BUS domain again disadvantages of a or b (400/500) and costs of a or b (406/506) were of interest in both tasks. For the remaining (frequent) items the two tasks differed quite substantially. In the drilling task consequences of a or b (700/701) reached 50% support. The information about probability of "everything is in order" (200/300) and the

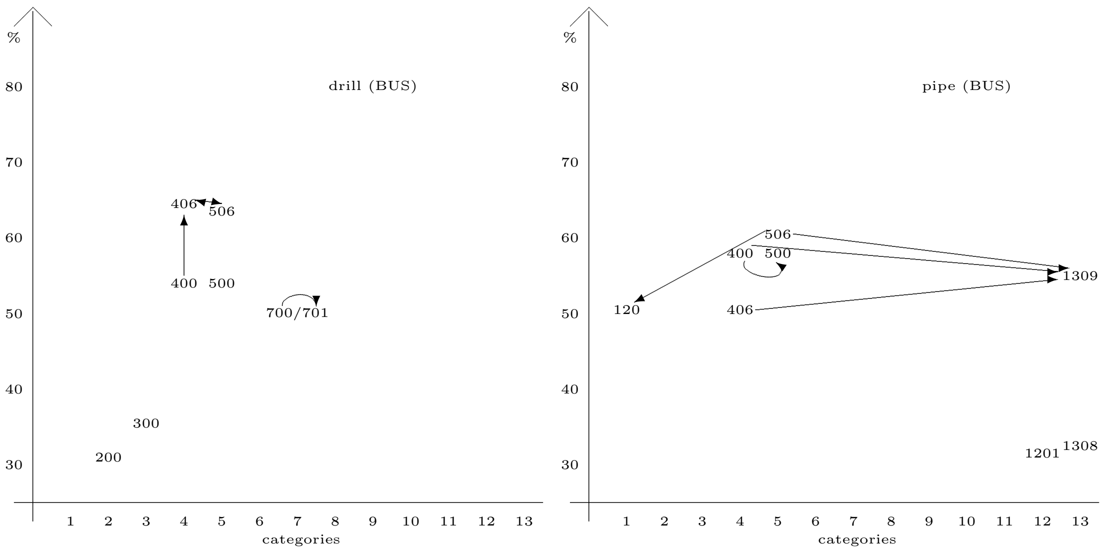


Fig. 32. Patterns for the drill and pipe task (BUS). Arrows indicate a sequence from item $x \rightarrow y$

sequences of costs of *a* and *b* (406 506) as well as disadvantages and costs of option *a* (400 406) were of interest, too.

In the pipe task a different picture emerged, here the question about costs of *b1* (1309) and the question about the "location of the maintenance" (120) both received support of more than %50. Furthermore two sequences were frequent the disadvantages question (400 500) and the costs of *b* followed by the question of "location of the maintenance" of option *a*. The basic question about what option *a* comprises (1308) and of option *a*'s alternatives to inhibit further negative events (1201) were only of interest in this task, too.

A third look at sequences - actual sequences

In the following analysis not a general pattern but the actual sequence of accessed items is of interest. Sequential information is very tricky when it comes to an analysis in statistical terms. Generally for each participant there exist 99 possibilities for the first click, then again 99 for the second ... given that an item can be accessed more than once. The tree diagram (see Figure 29, p. 96) of possible pathes grows quickly with ...

$$ss = x^n \tag{5}$$

... possibilities (ss = searchspace, x = search depth, n = number of possibilities, i.e. 99). Especially in consumer research the problem of sequences in search data are a common problem. In this area new approaches try to model customers' clickstreams (the sequence of accessed information) and their shopping behavior, i.e., a prediction from what a customer looks for on a Web site to the finally bought (or not bought) item (see e.g., Bucklin & Sismeiro, 2003; A. L. Montgomery, Li, Srinivasan, & Liechty, in press). These approaches use *Markov Models* (Freedman, 1983) for their predictions which require certain preconditions in the data not met in the current experiment. Precondition one is the necessity of a relative large set of data, e.g., A. L. Montgomery et al. (in press) used data from the Barnes & Nobel Web site including 1160 participants with over 9000 requested information items. Precondition two is the independence of observed items. In the current experiment this is not fulfilled because basically all items are equally probable (see formula 5) but the search space of a participant gets smaller with longer information search. The probability of accessing an item again (after it has been viewed) considerably shrinks for the first, the second ... click. Participants only "come back" to an item when they want to check for numbers and facts or can not remember a value looked up before. Therefore precondition two is also not met within the current experiment. Still, e.g., (A. L. Montgomery et al., in press) ignores the independence criterion, too – the data from a search process in a bookstore are hardly independent as soon as a certain search depth is reached (a book already considered as not interesting will most probably not be inspected again). In psychological research the author could not identify any research using this approach, therefore another solution that accounts for position in the search was developed.

Table 26. Percentage of clicks per item for the first 10 clicks - MEDICINE

ID	1.[a]	2.	3.	4.	5.	6.	7.	8.	9.	10.	sum	ID	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	sum
DRUG												OP											
202					5.04						5.04	200									5.56		5.56
												201		7.28	5.04	8.09	5.43	6.35	7.56	8.77	7.41		55.93
												203	5.19			7.35	7.75	5.56	5.04		5.56	8.91	45.36
300		5.92									5.92	300										5.94	5.94
301	5.70										5.70	301	5.19				5.43	6.35	5.88	6.14	7.41	10.89	47.29
												303			5.76	8.09	7.75		5.04	5.26			31.90
400	9.49	26.97	7.43	6.21		9.70					59.81	400	22.08	6.62					6.72				35.42
404									5.04		5.04	405					9.30						9.30
												406	5.19	5.30	5.04		5.43	13.49		6.14		6.93	47.52
500	30.38	9.87	8.11	5.52	5.04						58.91	500	16.88					6.35	7.56				30.80
505									6.72		6.72	506			5.76	5.88	5.43	7.14		5.26	6.48		35.95
701						5.22					5.22	700			8.63	8.82			5.04		5.56		28.05
707			10.14								10.14	701							5.04	5.26	5.56		15.86
708				8.28							8.28												
709	5.70						6.02				11.71												
711										5.61	5.61												
712		5.92									5.92												
714	9.49										9.49												
1200										5.61	5.61	1200		7.95							5.56		13.50
1201								6.35			6.35	1202		6.62	5.04					7.89			19.55
1305		5.92		6.21							12.13	1305										6.93	6.93

[a] numbers in the headline indicate the search steps: 1^{st}, 2^{nd}, 3^{rd} ... item of interest

The percentage of selected items per click is used as a dependent variable. However certain restrictions must be applied to the current data set. Because of the relatively fast spread of the information search with an increase of clicks only the first 10 clicks are taken into account (additionally, for these clicks only those are listed that occur in more than 5% of the cases). It is important to note that with this method the individual sequences are lost in favor of a *most common* path (for the begin of the search process).

Table 26 lists the percentages of clicks per item for the two tasks (drug, op) from the MED domain. A high number indicates that the general interest in this item (the ID indicates the corresponding item) was large in the x^{th} search step ($x \leq 10$). In the drug task the "side effects" of option *b* (500) are of interest for more than 30% of participants in the first click. Furthermore the combination of *a* and *b* (714) and "side effects" of *a* (400) are chosen frequently. In the second click clearly the attention switches to the "side effects" of *a* (400), in the third click "other side effects" (707) get the highest attention.

In the operation task the first click is dominated by an interest in "side effects" of *a* and *b*. Already with the second click the interest spreads apart - "probability of healing" of option *a* (201) and a general question about options against the side effects (1200) receive the highest interest here. In click 5 and 6 the information about the "course of the therapy" of option *a* and costs of this option are of interest.

In the law domain (see Table 27) the jail task's first click concentrates on disadvantages of *a* and *b* (400/500) and switches in the second click to closer information about "troubles" with option *a* (709). For the third to sixth click no favorite information can be found whereas in the seventh click disadvantages of *a* (400) gain larger (9.01 %) interest again.

Table 27. Percentage of clicks per item for the first 10 clicks - LAW

ID	1.[a]	2.	3.	4.	5.	6.	7.	8.	9.	10.	sum
JAIL											
200						5.93				5.68	11.61
202					5.83			6.54	7.00		19.38
203									6.00		6.00
302								5.61			5.61
400	21.83				5.83	5.93	9.01			5.68	48.29
401						5.08			6.00		11.08
406				5.51			6.31				11.82
500	19.01	6.52						5.61			31.14
501			5.26								5.26
506		6.52	5.26					7.48			19.26
700			5.26				7.21	5.61		6.82	24.90
709		7.97									7.97
710			6.02								6.02
711				5.51							5.51
712										5.68	5.68
1200			6.02						6.00		12.02
1305				5.51							5.51

ID	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	sum
REHAB											
110		7.36					6.11		5.04		18.51
117	5.45										5.45
201							5.34				5.34
203						5.19					5.19
400	21.82		5.10	5.20	5.67	5.19			5.88		43.66
401										7.21	7.21
405							7.63				7.63
406						5.93	6.11	6.30			18.33
500	14.55	11.04								5.41	30.99
501								5.51			5.51
502		6.75									6.75
506						6.67		5.51			12.18
700						5.19	5.34				10.53
701	5.45				6.38	6.67		5.51		5.41	29.42

[a] numbers in the headline indicate the search steps: 1^{st}, 2^{nd}, 3^{rd} ... item of interest

For the first click in the rehabilitation task an interest for disadvantages for *a* and *b* (400/500) with the later one (500) remaining also in the second click emerges. For the rest of the first ten clicks no dominating item could be found.

Finally the business domain is depicted in Table 28 with the pipe and drilling task. It is noticeable that these two tasks showed very few information items exceeding 5% of usage. In the pipe task again the disadvantages items appear in the first click (400/500) together with item 703 asking for general costs for both companies. Also new in this task was the appearance of the basic question of “why the repair has to be done” (117) in the second click. In the forth click costs of option *a* (406) were of interest. A substantial part of information search was dedicated to items 1308 and 1309 which provided “general information” about *a1* and “costs” of *a1*.

In the drilling task disadvantages and costs were central in click one with 400/500 (disadvantage of *a* and *b*) and combined costs of *a* and *b* (702). Consequences of option *a* (701) in click two and from click three to six an interest in costs of option *b* (506) were found. This pattern was also found (however a little less pronounced) in option *a* (406).

Table 28. Percentage of clicks per item for the first 10 clicks - BUSINESS

ID	1.[a]	2.	3.	4.	5.	6.	7.	8.	9.	10.	sum
PIPE											
117		10.22									10.22
120			6.06	7.81	8.13	7.08		5.05			34.13
400	23.94	6.57								5.62	36.13
406		5.84		11.72							17.56
500	15.49	8.03	6.82								30.34
501										6.74	6.74
506		8.76	8.33	5.47		5.31			5.32		33.19
703	14.79		6.06								20.85
1308		6.57		5.47							12.04
1309			5.30		9.76	6.19	7.77	8.08			37.10

ID	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	sum
DRILL											
400	15.15					5.11					20.26
406	6.06	6.75	8.92		10.96						32.69
500	11.52	6.13							7.02		24.67
506		7.36	8.28	9.21	8.22	7.30		5.17	5.26		50.81
700	7.88	7.36	5.10	5.92							26.26
701		11.04			5.48		5.38			6.60	28.51
702	14.55		6.37								20.91
703		9.82									9.82
709							6.92				6.92
712										5.66	5.66

[a] numbers in the headline indicate the search steps: 1^{st}, 2^{nd}, 3^{rd} ... item of interest

The final analysis in this study picks up the issue of sequences from third perspective. In the above analyses only the most common sequences, the first 10 clicks or the transactions between categories were considered. Because the end of the search process varies between participants it is not possible to consider the last items for a comparable analysis. Therefore the following steps are

Table 29. Mean Ranks of Categories

	drug	op	drill	pipe	rehab	jail
small						
general Info		2.40	2.78	3.52	3.29	3.25
P alt a	[a] 4.73	3.45	3.88		5.23	4.42
P alt b	3.91	3.08	4.07		3.57	3.40
alt a	3.41	3.30	3.19	2.65	3.78	3.27
alt b	3.54	3.58	3.03	2.60	3.83	3.25
rdo						
cons	4.05	3.38	3.00	2.38	3.74	3.59
new alt						
general P						
P a1						
P a2						
a1	5.60	2.67	4.40		4.92	3.60
b1	4.74		5.08	3.94	3.75	3.29
medium						
general Info	8.68	6.47	6.14	6.57	6.96	6.06
P alt a	8.35	6.24	7.46	7.94	8.03	7.11
P alt b	6.31	6.85	6.96	8.07	7.19	6.37
alt a	7.80	5.87	6.13	4.97	6.69	6.04
alt b	6.84	6.87	6.15	6.25	6.59	6.10
rdo						
cons	7.34	5.81	6.29	6.00	7.93	6.24
new alt				4.20		
general P				5.90		
P a1				8.00		
P a2						
a1	8.56	5.78	9.00	6.68	7.05	6.75
b1	6.31	6.60	8.58	6.97	8.14	6.05
large						
general Info	13.71	11.37	14.07	12.00	14.58	12.51
P alt a	15.40	10.68	15.56	11.00	17.31	13.71
P alt b	14.65	11.72	17.43	12.96	15.72	14.90
alt a	12.99	11.14	13.30	11.42	15.20	10.76
alt b	11.99	11.09	15.80	11.96	14.52	12.13
rdo			13.00			
cons	11.08	12.30	13.07	11.94	16.13	12.65
new alt		10.25	15.72	11.48	16.55	12.80
general P			9.09	12.84	15.60	13.00
P a1		13.09	20.37	14.06	19.33	19.20
P a2			22.16	16.76	15.96	13.68
a1	13.85	12.04	15.16	11.78	18.20	13.19
b1	13.22	14.94	17.94	13.31	16.93	13.59

[a] the lower the rank the earlier was the category considered in the search process

suggested. As reported in Chapter 8.2.2, p. 83 participants were grouped into a *small*-, *medium*- and *large*-searchgroup. For these three groups Table 29 depicts the *mean ranks* for the 13 categories and 6 tasks per searchgroup. The smaller the numbers in this table the earlier was the category selected *on average* over all participants. The three searchgroups are important for this analysis because an aggregation including all search lengthes would cover most of the effects. As a second point it is interesting to see how search length of a group of participants influences the positions of certain information categories.

The blank cells in this table result in the case of a total access to the cell of less than 10 clicks. This reduction was done because of an overweighing in resulting means of cells with small access. One category – *rdo* – only appears in the largest search group in one task (drill) however in this task the achieved rank is relatively low, i.e., a popular/early usage of this information type was present. The following categories are not ranked in the *small* search group: *new alt, general P, P a1* and *P a2*; this is still true for all but the *pipe* task in the *medium* search group. In the *large* search group the picture changes to an usage of the whole range of categories in all but the *op* (exceptions in categories are *new alt* and *P a1*) and *drug* task (in this task again *new alt, general P, P a1* and *P a2* are not accessed).

What are on average the first and the last categories participants access? For the *small* search group in four out of six tasks the *general Info* category results in the smallest rank means, i.e., this category was accessed on average before the other categories. The categories with the highest ranks (i.e., the last accessed categories) are the same for *drill* and *pipe* – category *b1* and for *rehab* and *jail* – category *P alt a*. In the *drug* task category *a1* and in the *op* task category *alt b* are the highest ranked categories.

In the *medium* search group results are not that clear. In the *drug* and *op* the category *P alt a/b* is ranked first whereas in *rehab* and *jail* it is *alt a/b*. The *pipe* task is the only one in this search group starting off with *new alt* whereas participants in the *drill* task showed interest for *alt a/b* and *general Info* first. On the other "end" of the search sequence *rehab* and *jail* ranked last with the categories *a1/b1.* The other tasks showed no systematic coverage.

In the third – the *large* search group – low means were found for *rehab* and *jail* with *alt a/b*. The other tasks showed no similarities. For the categories with the largest means (high ranks) *drill, pipe, rehab* and *jail* all used the *P a1/b1* category. For *drug* and *op* again different categories (*P alt a* and *b1* respectively) emerged.

8.3 Discussion of the domain issue

Three questions were raised at the beginning of this experiment. The following conclusions about them can be drawn from the above described data.

Do two tasks ascribed to one domain (by the researcher) result in comparable interest in information, i.e., can they be combined to one domain?

The general practice in research on decision domains is the usage of one task as being representative for one domain. The critical test of the assumption that a *domain effect* is measured and not a *task effect* was done in this experiment through the creation of two tasks for one domain and the empirical check whether these two tasks can be combined. The analysis was performed on several levels – the following considerations are organized from the most general to the most detailed level.

Taking a look at mean clicks for the six tasks an increase in gathered information *op* < *pipe* < *jail* < *drug* < *drill* < *rehab* was found. **The requirement of no mean differences in accessed information on the task level between tasks from one domain (e.g., drug/op) to another domain (e.g., drill/pipe) could not be demonstrated.** Actually there seems to be two groups of tasks: a *low search* group with *op, pipe* and *jail* and a *high search* with *drug, drill* and *rehab*. Each of these two groups includes all three domains but with shifted means.

A combination of the tasks in the intended pairs did not result in significant differences between domains. But what does an increase in accessed information mean? On the one hand it shows that a task is the *more demanding* the more information is accessed. This assumption is drawn because especially in an online experiment participants are not bound to the social pressure given in an experimental setting at, e.g., an university laboratory (the effect of a laboratory setting was demonstrated in experiment 1 and 2 – Chapter 6, p. 53). A participant in an online experiment is free to stop a task as soon as the amount of information gathered is satisfying to answer a problem. On the other hand the *familiarity with a task* could have an influence on the general information need to answer it. The more familiar a situation is to a participant the less information is needed. The structural demand in, e.g., the *operation* task is therefore much smaller than in the *drilling* task – an adult participant should be aware of the most important structural components of a task like: choosing between two operations. However, most of the participants will have little insight into the business of drilling companies – a less familiar task environment.

Staying at the perspective of *familiarity* there are three groups concerning amount of demanded information: the lowest need of information is found with the *op* and the *pipe* task. A medium amount of information was demanded in the *drug* and *jail* task. The highest amount of demanded information resulted in the *drill* and *rehab* task. At least for the first and the third group the familiarity explanation holds. Operation is a medical task which is more or less common in everyday's life. The pipe task describes a situation where a mending of pipes is the topic – a large number of participants should be familiar with such a situation. Exactly the opposite is true for the drill and rehab tasks – no insights on structure should be available for these tasks. Still the "middle" two tasks (*drug, jail*) do not fit into this explanation where *drug* should be much more familiar (being a medical task) than the *jail* task (which should be near to the rehab task in terms of structural understanding).

It can be argued that the mean amount of clicks per task is a very crude level for answering a question of information interest. Additionally, the complexity of the tasks and the search process could demand a finer grained perspective. Therefore 13 categories and then a reduced set of six categories were compared in order to find differences between domains and/or tasks.

The overall comparison (without the inclusion of tasks, yet) resulted in a large interest on information about the *two alternatives* (*a/b*) and *consequences*. Participants started off with searches using the two alternatives (which were presented in all the task texts and therefore readily available. This result can also be found when the used keywords are considered – the two alternatives are entered in more than 50% of all tasks). The interest in consequences can not be explained from this perspective. However in Huber et al. (2001) similar results were obtained. For consequences with certain outcomes the authors reported requests for the consequences category between 65% and even 100%. Consequences therefore seem to be a central variable for participants – more important than, e.g., probabilities or new alternatives (a result that is in contrast to classical decision making theory).

This overall picture of interest was analyzed in more detail for the 13 categories. In an analysis of profiles concerning the mean amounts of clicks no differentiation between tasks and/or domains was possible. The general results described above are also true for the more detailed task level – *consequences* and information about alternatives *a/b* are of largest interest (an exception are the operation and pipe task (see familiarity argument above) where these three categories range among the others). Also for a reduced set of categories (it was suspected that through the merging of categories hidden effects could be detected) no differences were found in this analysis.

However, looking at pairwise comparisons between categories per tasks (see Figure 28, p. 90) some interesting patterns were found. Probability information differentiated between the pipe and the other five tasks. No need for probability in a business task with very small stakes (damage of pipes in one's house) in comparison to higher stakes in the other tasks could account for this result. A reversed picture, with highest interest in information about the second alternative's additional action *b1* in pipe contrary to the other tasks demonstrates that participant's information search turns to other objectives for this task.

The point I want to make here is, that despite the general usage of one task being representative for a domain, the data from this experiment clearly do not support such a practice. In an even more complex manner the data show a global search pattern (on the mean amount of clicks level) for all participants regardless of the condition (domain or task).

Does additional information in task texts (concerning hints on the structure of a task) change information search behavior or choices?

Three cue levels were introduced in this experiment. The *2 cues* condition presented the task and the two alternatives; the *3 cues* condition added a hint on possible additional problems; the *4 cues* condition gave the full structure of the experiment (see Figure 24, p. 80) to the participants. Given these three levels it was suspected that an increase in information about the task would result in an increase of accessed information (it was assumed that an increase in task length gives more information (potential keywords) to the participant and therefore enlarges their search space). This was **not** the case in the above experiment. **No difference were found between mean amount of accessed information in respect to the three cue levels. No differences were found concerning the choices made between the threes CUE levels.** Even on the level of entered keywords no clear trend is observable, despite an overall increase in entered keywords per CUE condition ($2cues < 3cues < 4cues$) within the single tasks only the MED tasks generated a similar pattern on this measure.

An interesting result concerning the different CUE levels is that in the *3 and 4 cues* condition certain parts of the tasks were not of interest to the participants at all. In the *drug* and *op* task the nomination for the *test* as a keyword was underneath 1%. This is especially interesting because a test in medical terms should be considered as normal and informative. Either it is enough to confront participants with the two alternatives of a medical problem and let them generate a search strategy on their own or additional information is simply ignored and the search is done without consideration of all the available information. This result also sheds light on a very basic question raised with the introduction of the AIS paradigm. Does the presentation of a ready made task text influence the decision makers behavior? Within the AIS paradigm it is common to let the participant search for information while giving away only short instructions. This is also true for the tasks used in this experiment. However, no effect could be found for this manipulation – participants seem to have a ready made schema for *how* to search for information. This schema is definitely *not* triggered by the amount of information presented to them. But how could we access such a schema? An interesting experiment could test what parts of the instruction (the task) participants can remember after one or two weeks. Through such a memory test a verification whether the additional information is ignored per se could be given. For the three CUE levels from this experiment this would mean constant (bad) memory performances within all three conditions. If such a test resulted in differences (the more task information given the more participants could remember) it would mean that the information is remembered, thus ignored in the actual search process.

Given the same structure and values in a decision task, does content change information search behavior?

Three different approaches for the analysis of patterns in the searches were introduced. The simple *transaction* between categories of accessed information is the most general, structural approach to the data. On the next level the *pattern* of information categories shows which items are of interest to a majority of participants. This analysis was followed by a comparison of the *actual sequences*.

The *transactions* analysis showed that participants tended to stay within a category for their information search. The dominant search strategy therefore was to start with either information about alternative *a* or *b*, exploit this category until no more sufficient results could be found and switch to another category then. Within this second category the same procedure was started - exploit of the available information until no more interesting insights were given. Taking a look at the structure of the the information presentation within WebDiP this behavior seems logical from an economic point of view (going back to the same keyword and looking for more results there is easier than coming up with new keywords again and again). However, from an experimental point of view this behavior results in problems when the actual question is: what information items/categories are of interest between tasks/domains.

Concerning the *patterns* for the MED tasks *costs* and *side effects* played an important role, despite the fact, that the keyword “costs” was not entered very often. Participants did access information about costs through entering one of the two alternatives as keywords and chose costs from the resulting list of questions. In the longer patterns (especially in the op task) clearly a *attribute-wise* strategy can be observed, e.g., after information about *chance of healing* for the two alternatives was gathered, the *costs* of the two alternatives were compared. When inspecting Figure 30, p. 97 it is easy to see that the patterns between the two tasks differ substantially (again a hint on differences between tasks, suspected to belong to the same domain, i.e., resulting in the same search pattern). All of the used categories in the op task are close together in the graph – which means that a concentration on few categories (in this case those are the central ones from the task structure – probability of a/b and information about alternative a/b) in a combination with high support (large overlay between categories) is given. A very different picture arises in the drug task. Here a wide spread between patterns points at differences in support on the one hand (the y-axis) and usage of a larger variation in categories (consequences and information about a subsequent action is added), on the other hand.

In the LAW domain very similar patterns emerged between tasks – *disadvantages, costs* and *consequences* of the two alternatives received the highest percentages and therefore were of interest to a majority of participants. The most frequent patterns in the two tasks were followed both an*alternative*- and *attribute-wise* search with no pronunciation of one search mode. Inspecting the patterns in Figure 31, p. 98 the jail task only resulted in three sequences, all within information about the *two alternatives*. The rehab task was much more spread apparat including also information about consequences and probability information.

The third domain – BUS – included the tasks drill and pipe. Information about *disadvantages, costs* and *consequences* was popular in these two tasks. In the drill task the first direct comparison between *consequences* was found, whereas in the pipe task more interest in *costs* (even in costs of the additional action *b1*) was present. Again no clear differentiation between an *alternative*- and *attribute-wise* search was possible. The patterns in Figure 32, p. 99 illustrate a very "flat" picture for the pipe task, i.e., all sequences had considerably large support. Contrary to this pattern the drilling task is reduced to few sequences about costs (alternative a/b) and side effects (alternative a/b).

For the *actual sequences* the beginning of the search process (the first 10 items) and an overall ranking of categories. On the first click clearly a dominance of *disadvantages* (of the two alternatives) was found. Losses seem to loom larger when a search is started. Participants wanted to check whether a clearly negative alternative is available at first. Given that such an alternative would be found the decision process could be reduced to a minimum. This behavior fits into the predictions of SDS theory where an alternative is chosen in respect to its dominance over other alternative(s). In the two tasks from the MED domain participants continued with *additional probability* information about the *alternatives* and *consequences*. In the LAW domain possible *troubles* with the alternatives was of interest in the jail task whereas strictly *disadvantages* of *a* and *b* were observed in the rehab task. The clearest pattern was found in the BUS domain where in both tasks concentration was set on *costs* first. Even with a reduction of the range of clicks for the analysis to the first 10 it is hard (if not impossible) to identify differences between the tasks concerning used items. Quickly (starting with the second click) participant's searches spread to very broad search strategies using a multitude of items.

Stepping back from the level of items to the level of categories again. The summing up of the ranks of accessed categories is a tribute to the problem of overwhelming possibilities when the information search is viewed on the item level. Averaging over categories demands the splitting of the data into groups concerning *search depth*. This step introduces a new grouping variable not used in the analysis before. Additionally to an insight on what information categories are accessed on *average* at the beginning of searches also the end of the search sequences are accessible now. In the *small* search group tasks from the BUS and the LAW domain both ended on average with the same categories – *b1* and *P alt a* respectively (it is important to note that the average does not tell us the actual last category, but gives an impression of what categories are consulted relatively later in the search process than others). While participants in the BUS domain concentrate on *additional actions* of an alternative in the LAW domain still *probability* information is of interest. The two medical tasks showed no common *last* category.

Summing up, three information types gained the largest interest: *costs, consequences* and *disadvantages*. Participants concentrated on these three with only marginal deviations. The results for the MED domain show similarities in interest in the information categories *costs and side effects*. Additionally an *attribute-wise* search pattern was present in both tasks from this domain. Concerning search strategies the other four tasks (from LAW and BUS) resulted in a mixture out of alternative- and attribute-wise search. Note that the patterns described here do not include an absolute position of the pattern in the search stream. **Content largely influences the percentage of frequent search patterns. However no systematic similarities between domains could be found. Furthermore the amount of actually different categories is small.**

A quite diverse picture emerges when the above analyses are considered. Clearly different tasks use quite similar information categories an issue that will be addressed again in the general discussion below. This might account for the small differences between tasks when the mean amount of accessed information is considered. On the more detailed level of *patterns* and *sequences* differences between tasks arise (as expected), however, between domain – the main target of this experiment – no satisfactory similarities could be found. This is also true for the third reported analysis considering different cue levels – no effect could be demonstrated by changing the task texts considerable (in terms of presented information).

9

General Discussion

At the begin of this thesis two themes were introduced which served as a guide through the theory and experiments. These two themes were a methodological and a decision making one. The *methodological theme* dealt with two questions. Question 1 concerned the problem of how to run a process tracing study with a computer. Question 2 addressed the transfer from running experiments in the laboratory to Web-based studies.

The *decision making theme* highlighted the influence of content on the information search process, the patterns within this process and possible manipulations through the tasks' instructions. Relating to these two themes, three research questions were raised, which will be discussed first.

Transfer to the computer and comparison between Web and laboratory.

In experiment 1 and 2 the transfer from an AIS experiment into a computerized information search task was described. The AIS paradigm lends itself to be used in a computerized version because on an abstract level the method already includes all the steps that are performed in an information search task. The experimenter is replaced by a computer while the participant is working on a task which is very common in our daily lives, information search.

Further on, the second step in the transfer process was the introduction of an online information search experiment. In order to show that the transfer to the Internet has no effect on the results of the experiment at first a laboratory study was compared to the same study run on the Web. Results from this experiment demonstrated the basic applicability of a laboratory experiment on the Internet. Thus hints were found that technical problems in the search process accounted for differences in amount of searched information. However, the patterns of searches (the categories/types of accessed information) were the same in both locations. Additionally, arguments concerning the differences in samples when an Internet experiment is run could be disproved through the usage of a "real" Web sample and a comparison of this to students in both locations (laboratory and Web).

Technically the central problem in the earlier versions of the program were elicited through the usage of a small database for the search. Sufficient search results could not be guaranteed for all participants. This circumstance led to a large number of dropouts because most likely participants were frustrated after entering many keywords with no satisfying results. This was the starting point for WebDiP which represents a new approach for running information search experiments on the Internet. Through the usage of a keyword-guided search and a subsequent listing of possible "hits" participants find themselves in an environment they are used to nowadays. Large search engines like Google or Altavista use a similar system for finding information. It could be shown that an increase in accessed information was evident with the usage of WebDiP in comparison to the old system.

With the introduction of a new research tool a large number of questions is raised concerning the validity of the new tool in comparison with old ones. Researchers (especially cognitive psychologists) tend to distrust a method that can not be controlled the same way an experiment is controlled in the laboratory. There is no doubt that control in an experiment is of great importance and that experiments run on the Internet lack some of these control mechanisms. However a large number of methods is available which help to guarantee a certain amount of control on several dimensions. Examples for such control mechanisms are user registration via e-mail or check of IP-addresses to make it difficult participating in an experiment more than once. The multiple-site entry method (see Chapter 6.2.1, p. 55 for details) provides insight into sampling errors which help to find flaws connected to the population sampled from. Another example is the analysis of dropout behavior. The fact that no experimenter is present in an online experiment lets participants easily end an experiment and dropout. On the positive side, the analysis of dropouts can give insight into problems with a certain condition or task in an experiment. These can be corrected after pre-experiments or for subsequent experiments. Dropout hardly ever happens in a laboratory experiment therefore the possibility of easily finding problems here is not given. Using these and others control mechanisms for a large number of experiments a sufficient control level can be reached.

Assuming that somebody is interested in running a study online another often raised concern is that without technical knowledge and certain programming skills it is not possible to do so. WebDiP is a software example that provides the possibility to build an online information search study completely with the usage of a Web based interface. No editor, HTML or other technical knowledge is necessary. Even lay persons (students) were able to setup and run an experiment without extensive instructions. Taking away the pressure of technical boundaries of a tool or research method should increase willingness to use it and enlarge the number of experiments/studies done on the Internet.

Further advantages arose with the new system. WebDiP was released on sourceforge.net – a platform that provides access to a multitude of open source software. What does it mean for a research process if the tools a researcher is working with are freely available? Given WebDiP as an example, this means the whole program or parts of the code may be modified and also used in other software products, if they themselves are again released under the GPL (General Public License). Anybody who has an interest in the suggested method of information search can access the sourceforge.net Web site, download the program, install it on a computer and test whether the program fits the new research ideas. Most often flaws appear when a new research idea is used. In this case open source software guarantees that the researcher who is interested in additions or changes can start where the last developer stopped. Progress in research becomes considerably faster with such an approach. The current thesis not only introduced a new open source program but made also use of already available ones – the SPAM program used in *experiment 4* was also released at sourceforge.net and gave the author the possibility to shed light on otherwise not accessible patterns in the data stream produced by the participants. In a recent paper Vaughan (2004) illustrates that the Psychonomic Society introduced an online archive for data, stimuli and source code in 2004. The aim of this archive is to provide researchers with material used in published articles in one of Psychonomic's journals. This approach is in line with the above discussed open source idea and overdue for Psychology.

Several aims were outlined before the WebDiP project was started. From the author's perspective all of them have been reached within the current WebDiP development stage. In a test installation 15 experiments were run independently by 15 researchers (in this case students). No problems were reported concerning usability of the interface or data integrity between the experiments. The export of the generated data was performed smoothly either to a spreadsheet or even directly into SPSS. On the participants' side several studies already tested about 500 participants.

The only technical problems reported, concerned the relatively long URLs used within the e-mails sent to participants. Some mail-programs seem to have difficulties handling URLs longer than one line.

Further developments will include (1) enlarging the language support of the base system through systematic generation of as many language versions as possible; (2) generating of specific export filters from WebDiP into the SPAM and MAFIA algorithm; (3) introducing an open source experiment database. Through this database it should be possible to up- and download ready made experiments. Such a database would speed up the research cycle considerably, because task generation, which takes the most time within the experiment preparation, could be reduced to a minimum.

A second claim was made concerning the open source idea. Using open source development tools does not only provide up to date programming languages which are developed by a large number of programmers. Especially Web technologies like, e.g., the Apache server have extremely short patching times (i.e., correction of errors) when security flaws occur. The release of new products in the open source family considerably enlarges the group of available tools. The Web site www.sourceforge.net is an excellent example for the enormous number of free software available in the open source community – in November 2004 over 90,000 projects were listed – among them such important ones as PHP, Apache, PHPMyAdmin or Moodle (see Schulte-Mecklenbeck, 2004 for details). WebDiP might never be a big player on sourceforge.net but it already provides an interesting alternative to standard paper and pencil decision making tasks.

Decision making domains.

The decision making theme will be discussed next. As already discussed in Chapter 5, p. 41 decision *modes* are hard to elicit and even harder to keep constant (which is necessary for an experiment in the introduced paradigm). Therefore decision *domains* were used as the main dependent variable. Additional criticism was provided on a drawback in the current methodology which uses only one task as being representative for one decision domain. This confounding inhibits insight whether differences found within an experiment (between domains) are due to effects of the task or due to effects of the domain. This critical point was addressed in *Experiment 4* of this thesis. Concerning decision domains and the usage of two tasks for each of them the results of study 2 indicated that neither domains provided sufficient differences nor did the two tasks fit into one domain.

From various perspectives analyses were run on the three domain levels, each for two tasks from the (suspected) same domain. A first hint (from the perspective of the finished experiment) was already provided when the six tasks were generated. It was extremely difficult with certain information items to gain a sufficient homogeneity within two tasks from one domain or even more difficult between the six tasks overall. If the probability of a negative event is kept constant between two tasks content has a strong influence. A 5% chance of getting a disease is perceived as considerably higher than a 5% chance of loosing a contract with a company. Nevertheless, from a normative point of view these numbers are the same and should therefore result in the same decision behavior. Given the assumption that content changes a lot, why were no clearer patterns found between domains and tasks?

> "Identification of alternatives is important in real-life decisions but is often trivial in laboratory settings. In many real-life situations, decision problems appear with only one decision alternative at a time which has to be accepted or rejected."
> (Svenson, 1999, p. 178)

Svenson's quote is in many ways true and questions the validity of laboratory research at its very roots. Identification of alternatives was trivial in experiment 4 – alternatives were presented to the participants with the task texts. Participants made intense use of this "gift" – clearly on all reported measures the two alternatives were of central interest in terms of *magnitude* (how often was one of the alternatives *a* and *b* chosen as a keyword or as an information item), in terms of *keywords* and in terms of position in the *search sequence*. But is it a flaw or a feature?

As Blais and Weber (2001) demonstrated, different decision situations elicit the usage of specific reaction types or decision modes. It is obvious that a decision concerning a possible partner most often demands more emotional modi than a decision about money. In the current experiments another focus was set. The central interest was the type of information demanded by a participant in comparison between different tasks (domains). This is a much earlier stage (e.g., the third step in Svenson's model, Svenson, 1990) in the decision making process than the stage described by Blais and Weber (2001) where an actual decision is made. Furthermore in Blais and Weber (2001) as well as the other research approaches using decision modi, only one task is used as a representative for a domain. The current data do not support such a practice (the same behavior should have resulted for two tasks from one domain) quite the contrary was the case. These data argue for close inspection of experiments using only one task per domain. Despite a large effort in keeping everything constant between tasks/domains clearly the MED and BUS (at least pipe) domains resulted in a more familiar decision situation than the two tasks from the LAW domain.

Using WebDiP for finding domain differences

Before this project was started the author inspected available process tracing methods, especially interesting seemed the MouseLab approach introduced within the adaptive decision maker framework. A major critique that could be raised against this computerized information board approach is the following. The participant gets a lot of information about the task simply by inspecting the board and the there presented alternatives and attributes. So why not take this "bonus" from participants and bring them back to a much more naturalistic situation in which no structure is given? In real life situations the structured presentation of alternatives and attributes is only seldom available: consumer reports are one of the few exceptions. The WebDiP approach (as well as the AIS approach) does not provide any structural information but the two alternatives presented within the task text. Unfortunately these two alternatives have such a strong effect that a lot of the suspected differences could be covered by them. This is also true when the amount of information provided in the task is varied. No effect was found for this manipulation, too. Two modifications to the experiment could help overcoming these problems. The first modification is an experimental one that cuts down the number of information items a participant can access. Being allowed to used only a restricted number of information should increase the focus on the real important information and take away a lot of noise which is generated through the additional (less central) information items. A second modification aims at the simple clicking through the results of one keyword - a randomized results list would force participants to read all the resulting questions before choosing on of interest. Both modifications could be incorporated into future versions of WebDiP.

Another point of criticism concerning the information board method was the huge reductionism of the search process to alternative- or attribute-wise moves. The introduced *Payne Index* delivered one digit as a representative for the whole process (the Payne Index is questionable in other ways, too – see, e.g. Stokmans, 1992 discussion on ignorance of the "no transition" case and the "diagonal transition" case in Payne's index). However, from the perspective of a finished experiment which wanted to analyze the sequences in depth an index like the just described one would solve a multitude of problems. Even with the ignorance of a lot of processes in the search an experiment with a large number of participants, in many conditions results in an extremely difficulty to handle amount of different sequences. These sequences are hardly describeable without the use of computerized tools like MAFIA or SPAM. But even with the power of such tools common patterns are rare despite the above introduced critique concerning presentation of alternatives. The sheer statistical power of, e.g., Markov Models could cover a lot of the above described problems by including the sequentiality of data within the analysis. Regrettably the application of these models could not be performed within the current approach.

Payne et al. (1993) are not only well known for their methodological approach but also for the theoretical framework presented in the adaptive decision maker approach theory. Within this framework a tradeoff is described between *accuracy* and *effort* a decision maker wants to invest in a decision. The approach is appreciated because of the step from demonstrations of irrationality in the form of heuristics and biases to an understanding of the behavioral mechanisms underlying behavior. Within the WebDiP approach such a causal conclusion can not be made yet. Without the development of further tools for analysis and the generation of an index describing the sequentiality of the data (see below) causal mechanisms are hard to grasp. However, concerning effort a decision maker wants to invest the ADM theory clearly states that a tradeoff is performed. In the current study considerably long search sequences for a large number of participants clearly speak against such a tradeoff. Participants reported in their feedback that they often recognized the two alternatives as being the same. Why should they continue with their search after this insight? From an accuracy-effort point of view this behavior clearly violates the ADM predictions.

Developments

The question of *future developments* should be raised as the final one within this thesis. Because the technical issues were already discussed above only decision making issues will be mentioned here.

In experiment 4 the tasks' alternatives were balanced due to the following reasons. In pre-experiments (using the older software version) it turned out, that the search process is often terminated very early if an alternative outweighs another substantially (a finding described by Montgomery's search for a dominance structure). But through the balancing of alternatives also the opposite effect could result. Participants could stop their search (before they actually want to) because they are frustrated that no differentiating information about the alternatives can be located. Another strategy resulting from the balancing could be the extension of the search process, despite a satisfying result could be reached much earlier. Considering these two arguments and the fact, that the search sequences reached considerable lengths within WebDiP an experiment using alternatives that a clearly different seems intriguing. Not only would such a task be more naturalistic, but also a clearer stopping rule could be implied from the search processes. Stopping rules for a search process allow for implications about the end of a search not available in other settings (see, e.g., Browne & Pitts, 2004; Saad & Russo, 1996).

Additionally to this change in the task settings the development of an index inspired by Payne's Index would enable a researcher, using WebDiP, to compare tasks on a common denominator. However, the above discussed reduction of the search process to just one digit is not intended. One suggestion could be the usage of a *similarity matrix* which results in a score representing differences between sequences. This approach, originating from bio-informatics, compares sequences of binary data (the data from a search process can easily be represented as binary data using a hit/no-hit criterion for every information item) and provides similarity scores for the sequences. These scores would represent a more elegant version of the pattern and sequence approaches introduced in this thesis. A different approach would identify *target items* (items, which are of central interest to the researcher) first, and calculate the usage of these target items for each task then. This step would also provide a more comparable measure between tasks and domains.

Domain differences cover important areas of the psychological research field. However, in decision making research the question of domain differences was raised only recently in a broader sense. This thesis can contribute to the discussion on domain differences by demonstrating that two easy manipulations (adding a second task and varying information in it) can generate a considerable amount of mixed results which are not easily integrateable into existing approaches to the topic.

Researchers should recheck for effects actually not from the domain but from the selection of a task, which is especially generated for an experiment. Using at least this advice should result in clearer interpretable experiments using different domains.

> "Not only do people use different strategies for the same task, but a given individual may apply different strategies to subtly different tasks." (Payne et al. (1993), p. 70)

Appendix

A.1 Decision domain descriptions

Table 30: Domain Description for USort

Item	Shortform
what is objectively the best alternative	logical
no thought is given to a frequently encountered decision with importance	no thought
personal feeling that option is right	intuitive
choice is based on subjective preferences and feelings	emotional
postponing the decision	hesitant
choice according to other people's expectations	compliant
cost and benefits are considered; compromises and tradeoffs between the two are used for the decision	compromise mode
advice of respected other is considered	authority mode
avoiding of any action and remaining at the status-quo	avoidance mode
generating pros and cons; selecting overall best option	rational mode
intuition, emotions and gut feelings are key components	emotional
repeated and routinized situations	nondeliberative
following a plan for making the decision	rule-based
construct model; generate choice alternatives; when they fit, make decision	schema based
constructing stories/schemas	social
weighing advantages and disadvantages	inanimate objects
recognizing and classifying a situation; reacting in a "typical" way; if there is time: simulating outcome	recognition primed
abstract data; action that has to be justified	analytic
low in awareness and cognitive control	intuitive
automatic decision, based on expert knowledge	automatic
high cognitive control, high involvement and cognitive analysis	analytic
application of general rules, e.g., always choose the cheapest	rule based
application of context specific rules, e.g., choose the alternative most similar to situation X	rule based
after an action is set the outcome defines the ongoing decision	action based
the higher the sentimental value the less willingness to sell something	decisions about sentimental values
the more favorable a good is the less willingness to trade it	goal attribution
calculate outcome to decide	numerical calculation
identify worst outcome; pick option with best chance to avoid it	security
identify best outcome; pick option with best chance to get it	high aspiration
if regret happens; which option will I regret least	regret focused
simulation of feelings after choice	emotion focused
focus on morality;choose option which is "right"	morality focused
construct story for each option; pick best one	construct stories
which college shall I attend, which job shall I take	school decision
friendship related topics	relationship decision
whether to cry or not	group involving
if others do it - I try it too, e.g., drug use	group involving
whether to buy certain things or not	money related
choice among bet options	deep structure
content of decision is important to the choice	surface structure
weigh and combine the likelihood and desirability of outcomes	cost benefit
recognize a situation; apply a stored rule	rule based

parental decisions for their children (strengthening social identity)	role based
remember a situation in the past; take action	case based
affective reaction to different choice alternatives	affect based
if there is a red light; stop	nondeliberative
retrieval of stored judgment (e.g., opinion on a group)	stereotype based
if offered some illegal drugs; just say no	principle based
construction of alternative stories; evaluating of what might happen	schema based
seat-belt usage	health
skydiving versus bowling versus taking a walk	recreational
whether to confront co-workers with problems	social
whether to cheat on exams	ethical
identify options and chances of outcomes; assess values	analytic
build up collection of rules; match current situation to one in rule store	rule based
automatic matching of situation with rules	automatic

A.2 Instructions for USort

Step 1

The objective of Step 1 is to create logical groups from the items in the Source field. These items describe decision situations or methods/styles helping to make a decision.

1. To create each new group, drag an item from the Source field (on the left) into the Target field (on the right). The single line above and below items in the TARGET field shows that a group has been formed.
2. To add an item to an existing group, drag it from the Source field onto the line above the desired Target group. It will join the group when you release the mouse button. To move multiple items to a group, first click the checkbox to the left of each item, then drag the items simultaneously to the Target field. If you change your mind about where an item belongs, you can drag it to another group or back to the Source field. Leave the items you do not know in the Source field.
3. When you are satisfied with the groups you have created, click the right arrow to go to Step 2. Once you go to Step 2, you cannot return to Step 1.

Step 2

The objective of Step 2 is to combine the groups you have already created into higher-level groups. In other words, some of your groups may be similar and may belong together in "groups of groups".

To create the higher-level groups:

1. Move the similar groups next to each other by dragging the folder icon of one group atop the folder icon of the other.
2. Double-click the separator lines above and below the groups you have put together. The single lines will change to double-lines, which indicate a higher-level group.
   Repeat steps 1 and 2 for all higher-level groups.
3. When you have created all of your higher-level groups, check for any groups that have not been merged into higher-level groups. Change all the single lines between these original groups to double lines. You can double-click to change a line from single to double, and back again.

When you are satisfied with the higher-level groups you have created, click the right arrow to go to Step 3. Once you go to Step 3, you cannot return to Step 2.

Step 3

The objective of Step 3 is to give a descriptive name to each higher-level group of items.

To assign a name:

1. Click the double-line at the top of a group.
2. Type the name in the dialog box and click OK. Repeat for each group (double line).
3. When you have finished naming the groups, click the right arrow to finish your session. You will be prompted to save your input file.

A.3 Survey software and frequencies for domains

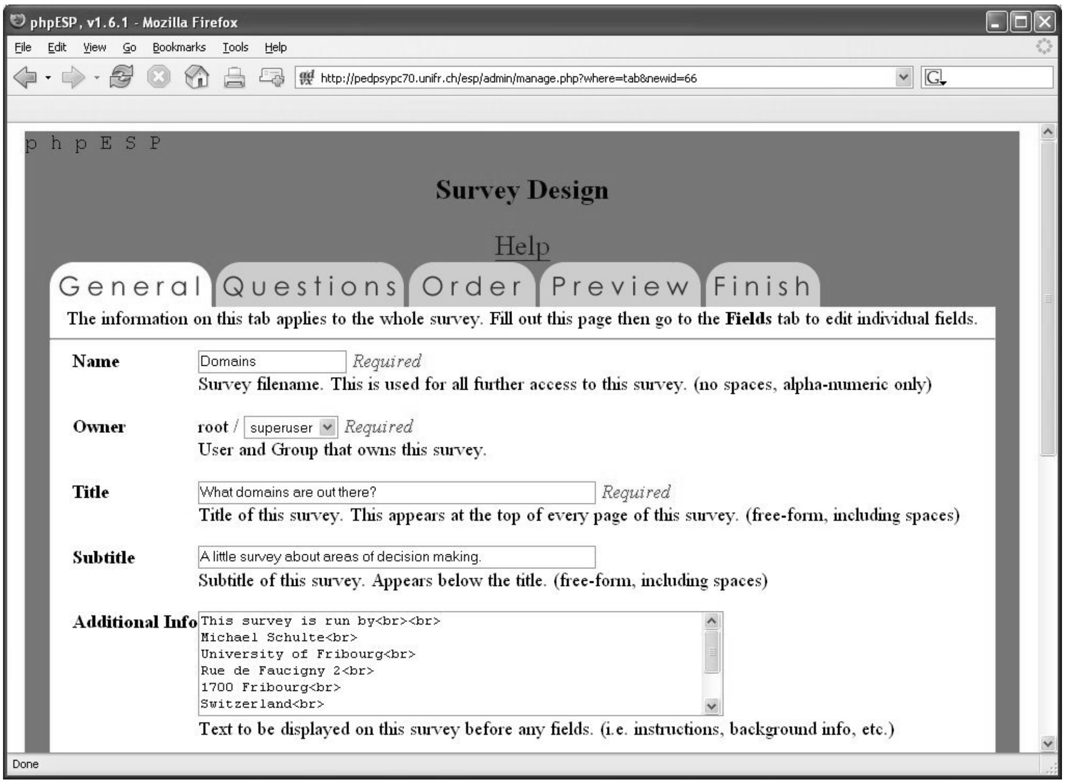


Fig. 33. Easy Survey Package Options

Table 31. Frequencies of domains

f^a	Domain
19	medical
14	law
13	business
12	politics
11	personal
10	health, finance
9	career, consumer
8	social, management
7	life, military
6	decision, gambling, marketing, relationships
5	children, family, marriage, organizational
4	public, science, weather
3	administration, auditing, buying, clinical, driving, economics, education, engineering, ethics, governmental, holiday, medicine, policy, purchasing, risk
2	accounting, agriculture, criminal, food, forecasting, friends, international, investment, money, police, professional, safety, sex, sports, strategic, work
1	academia, aerospace, aging, architecture, art, aviation, bargaining, behavior, biases, candidate, care, career, casino, choices, client, cognitive, communication, contraception, cooking, crisis, cultural, daily, decorating, development, diagnosis, diplomacy, disaster, drink, education, emergency, environment, environmental, estate, farming, genetic, goods, hazard, heuristics, home, impaired decision making, including, individual, investment, jobs, judgment, jury, landscaping, language, laypeople, leisure, literature, location, marriage, markets, mechanical, medicine, meteorology, music, negotiations, nutrition, perception, personnel, physics, poker, prediction, real, religious, research, Research, residence, routes, self-analysis, self-awareness, self-correction, selling, settlements, sociology, space, studying, style, technology, theater, traffic, transportation, travel, tv program, vocational, voter

[a] frequency of nomination(s)

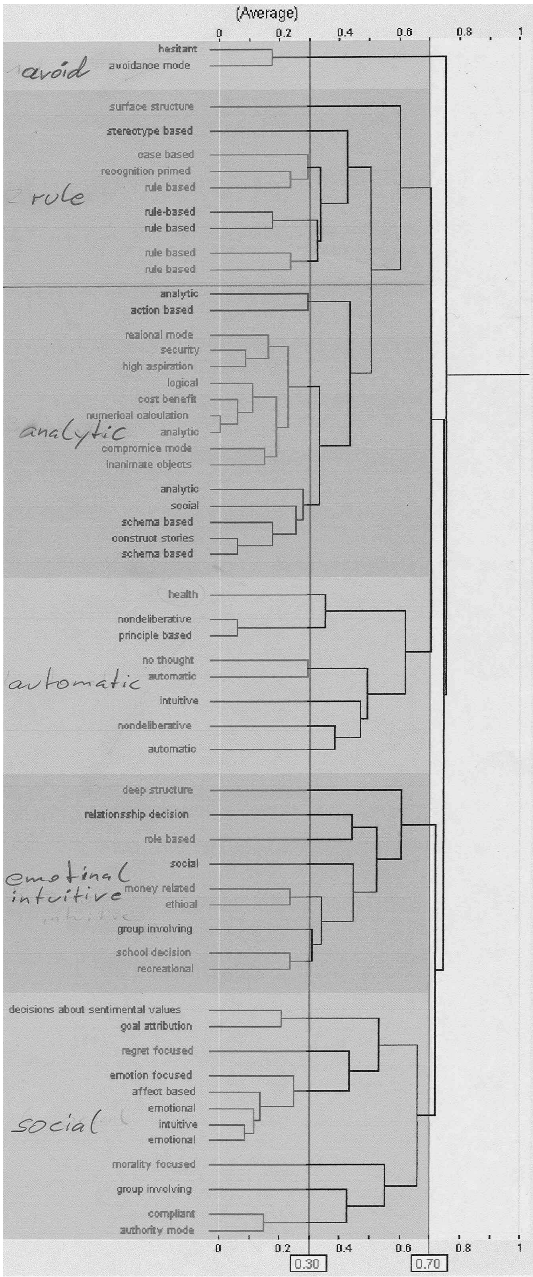


Fig. 34. Clusters resulting from EZCalc

A.4 Invitation e-mail for study 2

```
Liebe ListenteilnehmerInnen!

********************************************************************
Wie treffen wir Entscheidungen? Was für Informationen sind
wichtig, um eine Entscheidung zu treffen?
********************************************************************

Das Institut für Psychologie der Universität Freiburg (Schweiz)
führt ein wissenschaftliches Experiment zum Thema Entscheiden
durch.

Wenn Sie schon öfter wissen wollten wie Sie Entscheidungen treffen
ist dies eine gute Möglichkeit!

Das Experiment behandelt Informationen, die sie brauchen um eine
Aufgabe zu lösen. Die Dauer des Experimentes beträgt ca. 10
Minuten.

Als kleines Dankeschön werden unter den TeilnehmerInnen mit
abgeschlossenem Experiment 3x 50.- Euro  verlost.

Falls Sie Fragen oder Probleme haben, zögern Sie nicht mich unter
folgender e-mail Adresse zu kontaktieren: michael.schulte@unifr.ch

Viel Spass beim Experiment!

Bitte clicken Sie auf folgenden link um sich anzumelden.

http://pedpsypc70.unifr.ch/webdip/j22tsil.php

Bitte achten Sie darauf, dass der link vollständig in Ihren
Browser übernommen wird.


Mit freundlichen Grüssen

Michael Schulte
```

A.5 Task texts and questions for the tasks

This section provides the task texts for the six tasks. Because the experiment was run in German speaking countries (Austria, Germany and Switzerland) the original versions are provided. In the following section A.5.1 the questions for each task as provided for the participants are presented.

Business - Drilling

Task Text 2 cues

Stellen Sie sich vor, dass Sie ein Bauunternehmen leiten. Sie führen einen Auftrag durch bei dem mit einem Bohrroboter eine Röhre durch einen Berg gebohrt werden soll. Sie können sich zwischen zwei Bohrköpfen entscheiden: Bohrkopf der Firma Brosch oder Bohrkopf der Firma Holcim.

Task Text 3 cues

Stellen Sie sich vor, dass Sie ein Bauunternehmen leiten. Sie führen einen Auftrag durch bei dem mit einem Bohrroboter eine Röhre durch einen Berg gebohrt werden soll. Sie können sich zwischen zwei Bohrköpfen entscheiden: Bohrkopf der Firma Brosch oder Bohrkopf der Firma Holcim.

Es kann zu einer Zerstörung der Bohrköpfe kommen.

Task Text 4 cues

Stellen Sie sich vor, dass Sie ein Bauunternehmen leiten. Sie führen einen Auftrag durch bei dem mit einem Bohrroboter eine Röhre durch einen Berg gebohrt werden soll. Sie können sich zwischen zwei Bohrköpfen entscheiden: Bohrkopf der Firma Brosch oder Bohrkopf der Firma Holcim.

Gegen eine mögliche Zerstörung der Bohrköpfe stehen Zusatzgeräte zur Verfügung.

Business - Pipe

Task Text 2 cues

Stellen Sie sich vor, Sie besitzen ein Haus. Es gibt einen Wasserrohrbuch im Bad. Zur Reparatur stehen zwei Firmen zur Verfügung. Firma Bucher und Firma Steiner.

Task Text 3 cues

Stellen Sie sich vor, Sie besitzen ein Haus. Es gibt einen Wasserrohrbuch im Bad. Zur Reparatur stehen zwei Firmen zur Verfügung. Firma Bucher und Firma Steiner.

Es kann zu weiteren Wasserrohrbrüchen kommen.

Task Text 4 cues

Stellen Sie sich vor, Sie besitzen ein Haus. Es gibt einen Wasserrohrbuch im Bad. Zur Reparatur stehen zwei Firmen zur Verfügung. Firma Bucher und Firma Steiner.

Gegen mögliche weitere Wasserrohrbrüche stehen Zusatzpakete zur Verfügung.

Medicine - Drugs

Task Text 2 cues

Stellen Sie sich vor, Sie leiden seit einer gewissen Zeit unter starken Kopfschmerzen. Sie konsultieren einen Spezialisten, dieser diagnostiziert eine seltene Gehirnkrankheit. Zur Heilung existieren zwei Medikamente: Spinox und Lofa.

Task Text 3 cues

Stellen Sie sich vor, Sie leiden seit einer gewissen Zeit unter starken Kopfschmerzen. Sie konsultieren einen Spezialisten, dieser diagnostiziert eine seltene Gehirnkrankheit. Zur Heilung existieren zwei Medikamente Spinox und Lofa.

Beide Medikamente können Nebenwirkungen haben.

Task Text 4 cues

Stellen Sie sich vor, Sie leiden seit einer gewissen Zeit unter starken Kopfschmerzen. Sie konsultieren einen Spezialisten, dieser diagnostiziert eine seltene Gehirnkrankheit. Zur Heilung existieren zwei Medikamente Spinox und Lofa.

Gegen mögliche Nebenwirkungen der Medikamente gibt es Zusatzmedikamente.

Medicine - Operation

Task Text 2 cues

Stellen Sie sich vor, sie haben schon länger Magenschmerzen. Sie gehen zu einem Spezialisten der feststellt, dass ein Gewächs vorhanden ist, das entfernt werden muss. Es stehen für diese Operation 2 Spitäler zur Verfügung: das Inselspital und die Südklinik.

Task Text 3 cues

Stellen Sie sich vor, sie haben schon länger Magenschmerzen. Sie gehen zu einem Spezialisten der feststellt, dass ein Gewächs vorhanden ist, das entfernt werden muss. Es stehen für diese Operation 2 Spitäler zur Verfügung: das Inselspital und die Südklinik.

In beiden Spitälern können nach der Operation Blutungen auftreten.

Task Text 4 cues

Stellen Sie sich vor, sie haben schon länger Magenschmerzen. Sie gehen zu einem Spezialisten der feststellt, dass ein Gewächs vorhanden ist, das entfernt werden muss. Es stehen für diese Operation 2 Spitäler zur Verfügung: das Inselspital und die Südklinik.

Gegen mögliche Blutungen nach der Operation gibt es Zusatzuntersuchungen.

Law - Jail

Task Text 2 cues

Stellen Sie sich vor, Sie sind Richter in einem Prozess. Der Angeklagte wurde verurteilt und Sie sollen nun bestimmen, in welches Gefängnis er bis zur Einlieferung in ein Zielgefängnis überstellt wird. Sie können zwischen den Gefängnissen Hinterheim und Tierau wählen.

Task Text 3 cues

Stellen Sie sich vor, Sie sind Richter in einem Prozess. Der Angeklagte wurde verurteilt und Sie sollen nun bestimmen, in welches Gefängnis er bis zur Einlieferung in ein Zielgefängnis überstellt wird. Sie können zwischen den Gefängnissen Hinterheim und Tierau wählen.

In beiden Gefängnissen kann es zu einem Ausbruch kommen.

Task Text 4 cues

Stellen Sie sich vor, Sie sind Richter in einem Prozess. Der Angeklagte wurde verurteilt und Sie sollen nun bestimmen, in welches Gefängnis er bis zur Einlieferung in ein Zielgefängnis überstellt wird. Sie können zwischen den Gefängnissen Hinterheim und Tierau wählen.

In beiden Gefängnissen gibt es gegen einen möglichen Ausbruch Zusatzmassnahmen.

Law - Rehabilitation

Task Text 2 cues

Stellen Sie sich vor, Sie sind Richter in einem Prozess. Der Angeklagte wurde verurteilt und Sie sollen nun bestimmen, in welches Gefängnis er überstellt wird. Sie können zwischen den Gefängnissen Lenkheim und Seewald wählen.

Task Text 3 cues

Stellen Sie sich vor, Sie sind Richter in einem Prozess. Der Angeklagte wurde verurteilt und Sie sollen nun bestimmen, in welches Gefängnis er überstellt wird. Sie können zwischen den Gefängnissen Lenkheim und Seewald wählen.

Es kann zu einem Rückfall des Gefangenen kommen.

Task Text 4 cues

Stellen Sie sich vor, Sie sind Richter in einem Prozess. Der Angeklagte wurde verurteilt und Sie sollen nun bestimmen, in welches Gefängnis er überstellt wird. Sie können zwischen den Gefängnissen Lenkheim und Seewald wählen.

Gegen einen möglichen Rückfall des Gefangenen stehen in beiden Gefängnissen Zusatzmassnahmen zur Verfügung.

A.5.1 Questions for the six tasks

For the six tasks (machine, drug, law, operation, rehabilitation and pipe) the ID and corresponding question from the database are listed. Some questions appear more than once these are denoted by e.g. 402/404 (question 402, 403 and 404 are the same). However participants saw these only once but with an additional link telling them that there is “more information” available.

Table 32. IDs for operation

cat^a	ID	Question
gi	100/101	Welche Klinik ist besser?
	102	Welche Symptome gibt es?
	103	Welche Kliniken führen die Operationen durch?
	104	Welche Operation ist gefährlicher?
	105/106	Wie lange dauert die Behandlung für beide Operationen?
	107/108	Wie lange ist die Behandlungsdauer?
	109	Wie lange dauert die Behandlung?
	110	Können die Operationen kombiniert werden?
	111	Wann kann die Blutung auftreten?
	112	Wie gut sind die Operationen erprobt?
	113	Was ist das für eine Operation?
	114	Wie werden die Operationen durchgeführt?
	115	Kann es sein, dass die Operationen nur Symptome bekämpfen?
	116	Mit welchem Zeitrahmen muss man rechnen?
	117	Wie entsteht ein Gewächs im Magen?
	118/119	Was schlägt der Arzt vor?
	120	Wo wird die Behandlung durchgeführt?
	121	Was bedeutet Blutungen im Magen?
	122	Was bedeutet unangenehme Untersuchung?
	123	Wo entsteht die Blutung?
	124	Sind in beiden Krankenhäusern die Operationen sofort möglich?
	125	Was ist Ulcus ventriculi?
	126	Wer ist der Spezialist?
P_a	200/201	Wie gross sind die Heilungschancen (Wahrscheinlichkeit) bei der Operation im Inselspital?
	202/203	Wie gross ist die Chance (Wahrscheinlichkeit), dass nach der Operation im Inselspital Blutungen auftreten?
P_b	300/301	Wie gross sind die Heilungschancen (Wahrscheinlichkeit) bei der Operation in der Südklinik?
	302/303	Wie gross ist die Chance (Wahrscheinlichkeit), dass nach der Operation in der Südklinik Blutungen auftreten?
alt *a*	400	Gibt es Nebenwirkungen bei der Operation im Inselspital?
	401	Was passiert wenn es zu einer Blutung kommt?
	402/404	Ist die Durchführung einer Endoskopie nach der Operation gefahrlos?
	405	Wie ist der Ablauf der Behandlung im Inselspital?
	406	Wie viel kostet die Operation im Inselspital? (Kosten)
alt *b*	500	Gibt es Nebenwirkungen bei der Operation in der Südklinik?
	501	Was passiert wenn es zu einer Blutung kommt?
	502/504	Ist die Durchführung einer Magenspiegelung nach der Operation gefahrlos?
	505	Wie ist der Ablauf der Behandlung in der Südklinik?
	506	Wie viel kostet die Operation in der Südklinik? (Kosten)
rdo	600	Kann man die Untersuchung angenehmer machen?
	601	Was für eine Zusatzuntersuchung (Alternative) gegen eine Blutung steht zur Verfügung?
cons	700	Was sind die Konsequenzen der Operation im Inselspital?
	701	Was sind die Konsequenzen der Operation in der Südklinik?
	702/703	Wie hoch ist der Preis (Kosten) der Operationen?
	704	Gibt es weitere Bedingungen, die an die Behandlung geknüpft sind?
	705	Kann eine Magenblutung ohne Behandlung gefährlich sein?
	706	Wie unangenehm ist die Untersuchung?
	707	Gibt es noch andere Nebenwirkungen?
	708	Treten nach Beendigung der Behandlung andere Nebenwirkungen auf?
	709/711	Gibt es Nebenwirkungen bei der Durchführung einer Endoskopie nach einer Operation?
	712/714	Gibt es Nebenwirkungen bei der Durchführung einer Magenspiegelung nach einer Operation?
new alt	800	Gibt es noch eine andere Möglichkeit, die Blutung zu verhindern?
	801	Kann ich eine andere Behandlung durchführen lassen?
	802	Kann ich statt einer Operation Medikamente einnehmen?
	803	Gibt es noch andere Untersuchungen?
	804	Gibt es noch andere Möglichkeiten, das Gewächs zu entfernen?
general P	900/901	Wie gross sind die Heilungschancen (Wahrscheinlichkeit)?
	902/903	Wie hoch ist die Prävalenz für Gewächse im Magen?
P_a1	1000/1001	Wie gut wirkt eine Endoskopie gegen die Blutung?
P_b1	1100/1100	Wie gross ist die Chance (Wahrscheinlichkeit), dass der Test eine Blutung nicht erkennt?
	1102/1103	Wie gross ist die Chance (Wahrscheinlichkeit), dass der Test eine Blutung erkennt?
	1104/1105	Wie gross ist die Chance (Wahrscheinlichkeit), dass der Test eine Blutung anzeigt?
a1	1200	Was kann ich gegen die Blutung (Nebenwirkung) bei der Operation im Inselspital tun?
	1201/1202	Was kann ich gegen die Nebenwirkung durch die Operation im Inselspital tun?
	1203	Was ist eine Endoskopie?
	1204	Wie viel kostet die Endoskopie? (Kosten)
b1	1300/1301	Wie gut wirkt eine Magenspiegelung gegen die Blutung?
	1302/1303	Wie genau ist der Test?
	1304	Was macht der Test?
	1305/1306	Was kann ich gegen die Blutung (Nebenwirkung) bei der Operation in der Südklinik tun?
	1307	Was kann ich gegen die Nebenwirkung durch die Operation in der Südklinik tun?
	1308	Was ist eine Magenspiegelung?
	1309	Wieviel kostet die Magenspiegelung? (Kosten)
	1310	Innerhalb welcher Frist muss die Magenspiegelung durchgeführt werden?

[a] cat = categories, see Table 17, p. 87 for a description of the categories

Table 33. IDs for drug

ID[a]	Question
100/101	Welches Medikament wirkt besser?
102	Welche Symptome gibt es?
103	Welche Medikamente gibt es?
104	Welches Medikament ist schädlicher?
105/106	Wie lange dauert die Behandlung mit beiden Medikamenten?
107/108	Wie lange ist die Behandlungsdauer?
109	Wie lange dauert die Behandlung?
110	Warum können die Medikamente nicht miteinander kombiniert werden?
111	Wann kann die Lähmung auftreten?
112	Wie gut sind die Medikamente erforscht?
113	Was ist das für ein Medikament?
114	Wie werden die Medikamente angewendet?
115	Kann es sein, dass die Medikamente nur Symptome bekämpfen?
116	Mit welchem Zeitrahmen muss man rechnen?
117	Wie entsteht die Krankheit?
118/119	Was schlägt der Arzt vor?
120	Wo wird die Behandlung durchgeführt?
121	Was bedeutet Lähmung der Gesichtsmuskulatur?
122	Was bedeutet körperliche Probleme?
123	Wo entsteht die Lähmung im Gesicht?
124	Stehen die Medikamente sofort zur Verfügung?
125	Was ist Berioenzephalitis?
126	Wer ist der Spezialist?
200/201	Wie gross sind die Heilungschancen (Wahrscheinlichkeit) mit Spinox?
202/203	Wie gross ist die Chance (Wahrscheinlichkeit), dass nach der Behandlung mit Spinox Lähmungserscheinungen auftreten?
300/301	Wie gross sind die Heilungschancen (Wahrscheinlichkeit) mit Lofa?
302/303	Wie gross ist die Chance (Wahrscheinlichkeit), dass nach der Behandlung mit Lofa Nervenzellen geschädigt werden?
400	Gibt es Nebenwirkungen bei Spinox?
401	Was passiert wenn diese Nervenzellen durch Spinox zerstört werden?
402/404	Ist die Kombination von Spinox und Tremol gefahrlos?
405	Wie ist der Ablauf der Behandlung mit Spinox?
406	Wieviel kostet Spinox? (Kosten)
500	Gibt es Nebenwirkungen bei Lofa?
501	Was passiert wenn diese Nervenzellen durch Lofa zerstört werden?
502/504	Ist die Kombination von Lofa und Rigolin gefahrlos?
505	Wie ist der Ablauf der Behandlung mit Lofa?
506	Wie viel kostet Lofa? (Kosten)
600	Kann man gegen Kopfschmerzen und Übelkeit ein Medikament nehmen?
601	Was für ein Zusatzmedikament (Alternative) gegen eine Lähmung steht zur Verfügung?
700	Was sind die Konsequenzen vom Medikament Spinox?
701	Was sind die Konsequenzen vom Medikament Lofa?
702/703	Wie hoch ist der Preis (Kosten) der Medikamente?
704	Gibt es weitere Bedingungen, die an die Behandlung geknüpft sind?
705	Kann die Gesichtsmuskulatur nach der Behandlung wieder aktiviert werden?
706	Wie stark sind die Kopfschmerzen und die Übelkeit?
707	Gibt es noch andere Nebenwirkungen?
708	Treten nach Beendigung der Behandlung andere Nebenwirkungen auf?
709/711	Gibt es Nebenwirkungen bei der Kombination von Spinox und Tremol?
712/714	Gibt es Nebenwirkungen bei der Kombination von Lofa und Rigolin?
800	Gibt es noch eine andere Möglichkeit, die Lähmung zu verhindern?
801	Kann ich ein anderes Medikament einnehmen?
802	Kann ich eine Operation durchführen lassen?
803	Gibt es noch andere Untersuchungen?
804	Gibt es noch andere Möglichkeiten, die Krankheit zu behandeln?
900/901	Wie gross sind die Heilungschancen (Wahrscheinlichkeit)?
902/903	Wie hoch ist die Prävalenz der Krankheit?
1000/1001	Wie gut wirkt Tremol gegen die Lähmung?
1100/1100	Wie gross ist die Chance (W.), dass der Test eine entstehende Lähmungserscheinung nicht erkennt?
1102/1103	Wie gross ist die Chance (W.), dass der Test eine entstehende Lähmungserscheinung erkennt?
1104/1105	Wie gross ist die Chance (W.), dass der Test eine entstehende Lähmungserscheinung anzeigt?
1200	Was kann ich gegen die Lähmung (Nebenwirkung) durch Spinox tun?
1201/1202	Was kann ich gegen die Nebenwirkung von Spinox tun?
1203	Was ist Tremol?
1204	Wieviel kostet Tremol? (Kosten)
1300/1301	Wie gut wirkt Rigolin gegen die Lähmung?
1302/1303	Wie genau ist der Test?
1304	Was macht der Test?
1305/1306	Was kann ich gegen die Lähmung (Nebenwirkung) durch Lofa tun?
1307	Was kann ich gegen die Nebenwirkung von Lofa tun?
1308	Was ist Rigolin?
1309	Wieviel kostet Rigolin? (Kosten)
1310	Innerhalb welcher Frist muss das zweite Medikament eingenommen werden?

[a] Please refer to Table 32, p. 125 for a list of categories per question. IDs of questions correspond for all six tasks.

Table 34. IDs for rehabilitation

ID[a]	Question
100/101	Welches Gefängnis hat die bessere Resozialisierung?
102	Welche Straftaten hat der Häftling begangen?
103	Welche Gefängnisse gibt es?
104	Welches Gefängnis hat weniger Erfolg in Bezug auf die Resozialisierung?
105/106	Wie lange dauert die Unterbringung in den beiden Gefängnissen?
107/108	Wie lange bleibt der Verurteilte im Gefängnis?
109	Wie lang ist die Unterbringung?
110	Kann die Unterbringung in Lenkheim und Seewald kombiniert werden?
111	Wann kann eine neue Straftat stattfinden?
112	Wie gut sind die Erfahrungen mit den Gefängnissen?
113	Was ist das für eine Resozialisierungsmassnahme?
114	Wozu werden die Gefängnisse verwendet?
115	Kann es sein, dass die Gefängnisse gar nicht notwendig sind?
116	Wie lange dauert die Strafe?
117	Wieso gibt es eine Verurteilung?
118/119	Was schlagen die Kollegen vor?
120	Wo liegen die Gefängnisse?
121	Was bedeutet Mangel an Sozialarbeitern?
122	Was bedeuten weitere Formulare?
123	Wieso könnte der Gefangene rückfällig werden?
124	Stehen beide Gefängnisse sofort zur Verfügung?
126	Wer ist der Angeklagte?
200/201	Wie gross sind die Chancen (Wahrscheinlichkeit), dass die Resozialisierung in Lenkheim funktioniert?
202/203	Wie gross ist die Chance (Wahrscheinlichkeit), dass es nach Lenkheim zu neuen Straftaten kommt?
300/301	Wie gross sind die Chancen (Wahrscheinlichkeit), dass die Resozialisierung in Seewald funktioniert?
302/303	Wie gross ist die Chance (Wahrscheinlichkeit), dass es nach Seewald zu neuen Straftaten kommt?
400	Gibt es Nachteile bei der Wahl von Lenkheim?
401	Was passiert, wenn das Gefängnis Lenkheim nicht genug Sozialarbeiter hat?
402/404	Kann in Lenkheim das Sozialprogramm durchgeführt werden?
405	Was passiert bei der Unterbringung in Lenkheim?
406	Wie viel kostet die Unterbringung in Lenkheim (Kosten)?
500	Gibt es Nachteile bei der Wahl von Seewald?
501	Was passiert, wenn das Gefängnis Seewald nicht genug Sozialarbeiter hat?
502/504	Kann ich den Gefangenen in Lenkheim in einer gemeinnützigen Tätigkeit unterbringen?
505	Wie ist der Ablauf der Unterbringung in Seewald?
506	Wie viel kostet die Unterbringung in Seewald (Kosten)?
600	Kann man die Formulare umgehen?
601	Welche Resozialisierungsmassnahmen (Alterantiven) gegen einen Rückfall stehen zur Verfügung?
700	Was sind die Konsequenzen, wenn ich ihn ins Gefängnis Lenkheim schicke?
701	Was sind die Konsequenzen, wenn ich ihn ins Gefängnis Seewald schicke?
702/703	Wie hoch sind die Kosten für die Unterbringung in den Gefängnissen?
704	Gibt es weitere Bedingungen, die an die Unterbringung geknüpft sind?
705	Würde nach neuen Straftaten die Anzahl der Sozialarbeiter erhöht werden?
706	Wieviele Formulare müssen ausgefüllt werden?
707	Gibt es noch zusätzliche Formulare?
708	Sind nach dem Ende der Zusatzmassnahmen noch andere Probleme zu erwarten?
709/711	Können Probleme mit dem Sozialprogramm in Lenkheim auftreten?
712/714	Können Probleme mit der gemeinnützigen Tätigkeit in Seewald auftreten?
800	Gibt es noch eine andere Möglichkeit, neue Straftaten zu verhindern?
801	Stehen noch andere Gefängnisse zur Wahl?
802	Kann ich den Gefangenen in psychologische Betreuung geben?
803	Gibt es vor der Strafe noch andere Tests?
804	Gibt es noch andere Möglichkeiten nach dem Ende der Strafe?
900/901	Wie gross sind die Chancen (Wahrscheinlichkeit), dass die Gefängnisse eine Resozialisierung schaffen?
902/903	Wie gross ist die Chance (Wahrscheinlichkeit), dass andere Straftäter ähnlich verurteilt werden?
1000/1001	Wie hilfreich ist das Sozialprogramm gegen einen Rückfall?
1100/1100	Wie gross ist die Chance (Wahrscheinlichkeit), dass der Persönlichkeitstest eine neue Staftat nicht erkennt?
1102/1103	Wie gross ist die Chance (Wahrscheinlichkeit), dass der Persönlichkeitstest eine neue Staftat erkennt?
1104/1105	Wie gross ist die Chance (Wahrscheinlichkeit), dass der Persönlichkeitstest eine neue Straftat anzeigt?
1200	Was kann ich gegen neue Straftaten nach dem Gefängnis Lenkheim tun?
1201/1202	Was kann ich gegen neue Straftaten nach Lenkheim tun?
1203	Was passiert im Sozialprogramm?
1204	Wie viel kostet das Sozialprogramm zusätzlich (Kosten)?
1300/1301	Wie hilfreich ist die gemeinnützige Tätigkeit gegen einen Rückfall?
1302/1303	Wie genau ist der Persönlichkeitstest?
1304	Was testet der Persönlichkeitstest?
1305/1306	Was kann ich gegen neue Straftaten nach dem Gefängnis Seewald tun?
1307	Was kann ich gegen neue Straftaten nach Seewald tun?
1308	Was passiert bei der gemeinnützigen Tätigkeit?
1309	Wie viel kostet die gemeinnützige Tätigkeit (Betreuung) zusätzlich (Kosten)?
1310	Innerhalb welcher Frist beginnt die gemeinnützige Tätigkeit?

[a] Please refer to Table 32, p. 125 for a list of categories per question. IDs of questions correspond for all six tasks.

Table 35. IDs for jail

ID[a]	Question
100/101	Gibt es in Tierau noch andere Bewährungstests zur Beurteilung?
102	Welches Gefängnis ist sicherer?
103	Welche Straftaten hat der Gefangene begangen bzw. weswegen wurde er verurteilt (Urteil)?
104	Welche Gefängnisse gibt es?
105/106	Welches Gefängnis ist unsicherer in Bezug auf einen Ausbruch?
107/108	Wie lange dauert die Unterbringung in den beiden Gefängnissen?
109	Wie lange bleibt der Verurteilte im Gefängnis?
110	Kann ich für die Unterbringung beide Gefängnisse kombinieren?
111	Wie lang ist die Unterbringung?
112	Wann kann ein Ausbruch stattfinden?
113	Wie gut sind die Erfahrungen mit den Gefängnissen?
114	Was ist das für eine Zusatzmassnahme?
115	Wozu werden die Gefängnisse verwendet?
116	Kann es sein, dass die Gefängnisse gar nicht notwendig sind?
117	Wie lange dauert die Unterbringung?
118/119	Wieso gibt es eine Verurteilung?
120	Was schlagen die Kollegen vor?
121	Wo liegen die Gefängnisse?
122	Was bedeutet Personalmangel?
123	Was bedeuten weitere Formulare?
124	Wo kann man ausbrechen?
125	Stehen beide Gefängnisse sofort zur Verfügung?
126	Wer ist der Angeklagte?
200/201	Wie gross sind die Chancen (Wahrscheinlichkeit), dass in Hinterheim alles funktioniert?
202/203	Wie gross ist die Chance (Wahrscheinlichkeit), dass es in Hinterheim zu einem Ausbruch kommt?
300/301	Wie gross sind die Chancen (Wahrscheinlichkeit), dass in Tierau alles funktioniert?
302/303	Wie gross ist die Chance (Wahrscheinlichkeit), dass es in Tierau zu einem Ausbruch kommt?
400	Gibt es Nachteile bei der Wahl von Hinterheim?
401	Was passiert wenn das Gefängnis Hinterheim nicht genug Personal hat?
402/404	Kann ich den Gefangenen in Hinterheim in Einzelhaft unterbringen lassen?
405	Was passiert bei der Unterbringung in Hinterheim?
406	Wie viel kostet die Unterbringung in Hinterheim (Kosten)?
500	Gibt es Nachteile bei der Wahl von Tierau?
501	Was passiert, wenn das Gefängnis Tierau nicht genug Personal hat?
502/504	Kann ich den Gefangenen in Tierau im Hochsicherheitstrakt unterbringen lassen?
505	Wie ist der Ablauf der Unterbringung in Tierau?
506	Wie viel kostet die Unterbringung in Tierau (Kosten)?
600	Kann man die Formulare umgehen?
601	Welche Zusatzmassnahme (Alternative) gegen einen Ausbruch steht zur Verfügung?
700	Was sind die Konsequenzen, wenn ich ihn ins Gefängnis Hinterheim schicke?
701	Was sind die Konsequenzen, wenn ich ihn ins Gefängnis Tierau schicke?
702/703	Wie hoch sind die Kosten für die Unterbringung in den Gefängnissen?
704	Gibt es weitere Bedingungen, die an die Unterbringung geknüpft sind?
705	Würde nach einem Ausbruch das Personal erhöht werden?
706	Wie viele Formulare müssen ausgefüllt werden?
707	Gibt es noch zusätzliche Formulare?
708	Sind nach dem Ende der Unterbringung noch andere Probleme zu erwarten?
709/711	Können Probleme mit der Einzelhaft in Hinterheim auftreten?
712/714	Können Probleme mit dem Hochsicherheitstrakt in Tierau auftreten?
800	Gibt es noch eine andere Möglichkeit, einen Ausbruch zu verhindern?
801	Stehen noch andere Gefängnisse zur Wahl?
802	Kann ich den Gefangenen in psychologische Betreuung geben?
803	Gibt es vor der Inhaftierung noch andere Möglichkeiten von Tests?
804	Gibt es noch andere Möglichkeiten, die Zeit bis zum Zielgefängnis zu überbrücken?
900/901	Wie gross sind die Chancen (Wahrscheinlichkeit), dass die Gefängnisse die ganze Strafzeit garantieren können?
902/903	Wie gross ist die Chance (Wahrscheinlichkeit)?
1000/1001	Wie hilfreich ist die Einzelhaft gegen einen Ausbruch?
1100/1100	Wie gross ist die Chance (Wahrscheinlichkeit), dass der Bewährungstest einen geplanten Ausbruch nicht erkennt?
1102/1103	Wie gross ist die Chance (Wahrscheinlichkeit), dass der Bewährungstest einen geplanten Ausbruch erkennt?
1104/1105	Wie gross ist die Chance (Wahrscheinlichkeit), dass der Bewährungstest einen Ausbruch anzeigt, obwohl gar keiner geplant ist?
1200	Was kann ich gegen die Ausbruchsversuche im Gefängnis Hinterheim tun?
1201/1202	Was kann ich gegen einen Ausbruch aus Hinterheim tun?
1203	Was passiert in der Einzelhaft?
1204	Wie viel kostet die Einzelhaft zusätzlich (Kosten)?
1300/1301	Wie hilfreich ist der Hochsicherheitstrakt gegen einen Ausbruch?
1302/1303	Wie genau ist der Bewährungstest?
1304	Was testet der Bewährungstest?
1305/1306	Was kann ich gegen einen Ausbruch aus Tierau tun?
1307	Was kann ich gegen einen Ausbruch aus Tierau tun?
1308	Was passiert im Hochsicherheitstrakt?
1309	Wie viel kostet der Hochsicherheitstrakt zusätzlich (Kosten)?
1310	Innerhalb welcher Frist muss der Gefangene in den Hochsicherheitstrakt?

[a] Please refer to Table 32, p. 125 for a list of categories per question. IDs of questions correspond for all six tasks.

Table 36. IDs for pipe

ID[a]	Question
100/101	Welche Firma hat den besseren Ruf?
102	Welche Anzeichen für einen Wasserrohrbruch gibt es?
103	Welche Firmen stehen zur Verfügung?
104	Welche Firma arbeitet schlechter?
105/106	Wie lange dauern die Arbeiten?
107/108	Ist die Arbeitsdauer der Firmen immer gleich lang?
109	Wie gross ist die Arbeitsdauer?
110	Kann ich beide Firmen engagieren?
111	Wann kann ein weiterer Wasserrohrbruch auftreten?
112	Wie gut sind die Erfahrungen mit den beiden Firmen?
113	Was ist das für eine Reparatur?
114	Wie arbeiten die Firmen?
115	Kann es sein, dass nur der Schaden behoben wird ohne die Ursache zu beseitigen?
116	Mit welchem Zeitrahmen muss man rechnen?
117	Wieso muss die Reparatur durchgeführt werden?
118/119	Was schlagen die Freunde vor?
120	Wo findet die Reparatur statt?
121	Was passiert, wenn weitere Rohre defekt sind?
122	Was bedeutet eine größere Verschmutzung?
123	Wo kann es zu einem Wasserrohrbruch kommen?
124	Haben die beiden Firmen sofort für die Reparatur Zeit?
125	In welchem Haus findet die Reparatur statt?
126	Wie kann man den Hausbesitzer (Sie) beschreiben?
200/201	Wie gross sind die Chancen (Wahrscheinlichkeit), dass die Firma Bucher erfolgreich ist?
202/203	Wie gross ist die Wahrscheinlichkeit (Chance), dass bei einem weiteren Wasserrohrbruch die Arbeit der Firma Bucher umsonst war?
300/301	Wie gross sind die Chancen (Wahrscheinlichkeit), dass die Firma Steiner erfolgreich ist?
302/303	Wie gross ist die Wahrscheinlichkeit (Chance), dass bei einem weiteren Wasserrohrbruch die Arbeit der Firma Steiner umsonst war?
400	Gibt es Nachteile wenn ich die Firma Bucher beauftrage?
401	Was kann passieren, wenn die Firma Bucher nur die Standardreparatur durchführt?
402/404	Kann die Firma Bucher die Standardreparatur und die Bucher Komplettsanierung gleichzeitig durchführen?
405	Wie ist das Vorgehen bei der Standardreparatur durch die Firma Bucher?
406	Wieviel kostet die Standardreparatur durch die Firma Bucher? (Kosten)
500	Gibt es Nachteile wenn ich die Firma Steiner beauftrage?
501	Was kann passieren, wenn die Firma Steiner nur die Grundreparatur durchführt?
502/504	Kann die Firma Steiner die Grundreparatur und die Steiner Totalsanierung gleichzeitig durchführen?
505	Wie ist das Vorgehen bei der Grundreparatur durch die Firma Steiner?
506	Wieviel kostet die Grundreparatur durch die Firma Steiner? (Kosten)
600	Kann man etwas gegen den Schmutz tun?
601	Welche Zusatzpakete (Alternativen) gegen weitere Rohrbrüche stehen zur Verfügung?
700	Was ist die Konsequenz wenn ich den Auftrag der Firma Bucher gebe?
701	Was ist die Konsequenz wenn ich den Auftrag der Firma Steiner gebe?
702/703	Wie viel kostet die Reparaturen bei den beiden Firmen? (Kosten)
704	Gibt es weitere Bedingungen, die an die Reparatur geknüpft sind?
705	Kann ein Wasserrohrbruch ohne Reparatur gefährlich sein?
706	In welchem Ausmass wird Schmutz vorhanden sein?
707	Gibt es noch andere Komplikationen, die auftreten können?
708	Können nach dem Ende der Arbeiten noch weitere Probleme auftreten?
709/711	Können Probleme auftreten, wenn die Standardreparatur und die Bucher Komplettsanierung gleichzeitig durchgeführt werden?
712/714	Können Probleme auftreten, wenn die Grundreparatur und die Steiner Totalsanierung gleichzeitig durchgeführt werden?
800	Gibt es noch eine andere Möglichkeit, einem weiteren Wasserrohrbruch vorzubeugen (Vorbeugung)?
801	Kann ich andere Firmen beauftragen?
802	Gibt es noch andere Möglichkeiten den Schaden zu beheben?
803	Hat die Firma Steiner noch andere Geräte um die übrigen Rohre zu untersuchen?
804	Kann ich die Rohre auch anders reparieren?
900/901	Wie gross sind die Chancen (Wahrscheinlichkeit), dass die Reparatur erfolgreich ist?
902/903	Wie gross ist die Wahrscheinlichkeit (Chance) eines Wasserrohrbruchs?
1000/1001	Wie hilfreich ist eine Bucher Komplettsanierung bei einem weiteren Wasserrohrbruch?
1100/1100	Wie gross ist die Wahrscheinlichkeit (Chance), dass das Magnetometer defekte Rohre nicht erkennt?
1102/1103	Wie gross ist die Wahrscheinlichkeit (Chance), dass das Magnetometer defekte Rohre erkennt?
1104/1105	Wie gross ist die Wahrscheinlichkeit (Chance), dass das Magnetometer defekte Rohre anzeigt?
1200	Was kann man gegen mögliche weitere Wasserrohrbrüche tun, wenn ich die Firma Bucher beauftrage?
1201/1202	Was kann die Firma Bucher gegen die Schäden durch einen möglichen weiteren Wasserrohrbruch tun?
1203	Was ist die Bucher Komplettsanierung?
1204	Wie viel kostet die Bucher Komplettsanierung? (Kosten)
1300/1301	Wie hilfreich ist eine Steiner Totalsanierung bei einem weiteren Wasserrohrbruch?
1302/1303	Wie genau ist das Magnetometer?
1304	Was testet das Magnetometer?
1305/1306	Was kann ich gegen mögliche weitere Wasserrohrbrüche tun, wenn ich die Firma Steiner beauftrage?
1307	Was kann die Firma Steiner gegen die Schäden durch einen möglichen weiteren Wasserrohrbruch tun?
1308	Was ist die Steiner Totalsanierung?
1309	Wie viel kostet die Steiner Totalsanierung? (Kosten)
1310	Innerhalb welcher Frist muss die Steiner Totalsanierung durchgeführt werden?

[a] Please refer to Table 32, p. 125 for a list of categories per question. IDs of questions correspond for all six tasks.

Table 37. IDs for drilling

ID[a]	Question
100/101	Welcher Bohrkopf ist besser?
102	Welche Anzeichen für Probleme bei der Bohrung gibt es?
103	Welche Bohrköpfe gibt es?
104	Welcher Bohrkopf ist anfälliger auf Granit?
105/106	Wie lange kann man die Bohrköpfe verwenden?
107/108	Ist die Lebensdauer der Bohrköpfe immer gleich lang?
109	Wie gross ist die Lebensdauer?
110	Kann ich die Bohrköpfe kombinieren?
111	Wann kann Granit auftreten?
112	Wie gut sind die Erfahrungen mit den Bohrköpfen?
113	Was ist das für ein Bohrroboter?
114	Wie werden die Bohrköpfe verwendet?
115	Kann es sein, dass die Bohrköpfe auch ohne das Vorhandensein von Granit heiss laufen?
116	Mit welchem Zeitrahmen muss man rechnen?
117	Wieso muss gebohrt werden?
118/119	Was schlagen die Kollegen vor?
120	Wo ist der Berg?
121	Was passiert, wenn Granit auftritt?
122	Was bedeutet schnellerer Verbrauch?
123	Wo kann der Granit vorkommen?
124	Stehen die Bohrköpfe sofort zur Verfügung?
125	Was ist das für ein Berg, durch den gebohrt werden soll?
126	Wie kann man das Bauunternehmen beschreiben?
200/201	Wie gross sind die Chancen (Wahrscheinlichkeit), dass Brosch funktioniert?
202/203	Wie gross ist die Wahrscheinlichkeit (Chance), dass beim Auftreten von Granit Brosch zerstört wird?
300/301	Wie gross sind die Chancen (Wahrscheinlichkeit), dass Holcim funktioniert?
302/303	Wie gross ist die Wahrscheinlichkeit (Chance), dass beim Auftreten von Granit Holcim zerstört wird?
400	Gibt es Nachteile bei der Verwendung von Brosch?
401	Was passiert, wenn man mit Brosch in Granit schnell bohrt?
402/404	Ist die Kombination von Brosch und Varim möglich?
405	Wie ist das Vorgehen bei der Bohrung mit Brosch?
406	Wieviel kostet Brosch? (Kosten)
500	Gibt es Nachteile bei der Verwendung von Holcim?
501	Was passiert, wenn man mit Holcim in Granit schnell bohrt?
502/504	Ist die Kombination von Holcim und Sertec möglich?
505	Wie ist das Vorgehen bei der Bohrung mit Holcim?
506	Wieviel kostet Holcim? (Kosten)
600	Kann man etwas gegen den schnelleren Verbrauch tun?
601	Was für ein Zusatzgerät (Alternative) gegen eine Zerstörung des Bohrkopfes steht zur Verfügung?
700	Was ist die Konsequenz bei Verwendung des Bohrkopfes der Firma Brosch?
701	Was ist die Konsequenz bei Verwendung des Bohrkopfes der Firma Holcim?
702/703	Wie hoch ist der Preis der Bohrköpfe? (Kosten)
704	Gibt es weitere Bedingungen, die an die Bohrung geknüpft sind?
705	Kann man den Bohrkopf nach einer Beschädigung reparieren?
706	In welchem Ausmass werden die Kleinteile verbraucht?
707	Gibt es noch andere Dinge, die dem Bohrkopf gefährlich werden können?
708	Können nach dem Ende der Bohrung noch weitere Probleme auftreten?
709/711	Können Probleme bei der Kombination von Brosch und Varim auftreten?
712/714	Können Probleme bei der Kombination von Holcim und Sertec auftreten?
800	Gibt es noch eine andere Möglichkeit, den Granit zu durchbohren?
801	Kann ich andere Bohrköpfe kaufen?
802	Kann ich auch sprengen statt bohren?
803	Gibt es vor der Bohrung noch andere Möglichkeiten der Gesteinsanalyse?
804	Gibt es noch andere Möglichkeiten die Bohrung durchzuführen?
900/901	Wie gross sind die Chancen (Wahrscheinlichkeit), dass die Bohrköpfe funktionieren?
902/903	Wie gross ist die Wahrscheinlichkeit eines Bohrauftrages dieser Grössenordnung?
1000/1001	Wie hilfreich ist Varim bei hartem Gestein?
1100/1100	Wie gross ist die Wahrscheinlichkeit (Chance), dass der Sensor Granit nicht erkennt?
1102/1103	Wie gross ist die Wahrscheinlichkeit (Chance), dass der Sensor Granit erkennt?
1104/1105	Wie gross ist die Wahrscheinlichkeit (Chance), dass der Sensor Granit anzeigt?
1200	Was kann ich gegen den Granit tun wenn ich Brosch verwende?
1201/1202	Was kann ich gegen die zu hohe Geschwindigkeit bei der Verwendung von Brosch tun?
1203	Was ist Varim?
1204	Wieviel kostet Varim? (Kosten)
1300/1301	Wie hilfreich ist Sertec bei hartem Gestein?
1302/1303	Wie genau ist der Sensor?
1304	Was testet der Sensor?
1305/1306	Was kann ich gegen den Granit tun wenn ich Holcim verwende?
1307	Was kann ich gegen die zu hohe Geschwindigkeit bei der Verwendung von Holcim tun?
1308	Was ist Sertec?
1309	Wieviel kostet Sertec? (Kosten)
1310	Innerhalb welcher Frist muss das Zusatzgerät Sertec eingebaut werden?

[a] Please refer to Table 32, p. 125 for a list of categories per question. IDs of questions correspond for all six tasks.

A.6 Data preparation

The 368 participants who finished the experiment generated a log file with 41119 lines in which every action was recorded. A fraction of this log files is described now:

```
ID USER TEST TASK ACTION CHOICE
1  1358 69   216  krank      0
2  1358 69   216  13395      0
```

```
3  1358 69   216  13394     0
4  1358 69   216  lofa      0
5  1358 69   216  13396     0
6  1358 69   216  lofa      0
7  1358 69   216  spinox    0
8  1358 69   216  13383     0
9  1358 69   216  taskend   452
10 1358 69   217  tasktimer 0
```

The first column (ID) continually numbers the actions of participants. In the second column (USER) the identification number of the participant 1358 is recorded, who is in the TEST (or experiment) condition 69 (the experiments differed in task sequence and sequence of alternative presentation, all of these conditions were counterbalanced). The TASK column identifies one of the six tasks used in the experiment – in the case of 216 the drug task was done. The ACTION column includes the entered keywords (krank, lofa, spinox), the information items looked at (13395, 13394, 13396 and 13383) and a timer variable defining the end of a task (taskend) and the beginning of a new one (tasktimer). Finally in the last column of the log (the last action within a task) the choice – is recorded, for user 1358 this was 452 (Lofa). In the last line the tasktimer is set back to zero again and a new task (217) is started.

This raw log file was first processed in order to number the tasks (the sequence of task for each participant) and information items (the sequence of information items (clicks) for each participant in each task). Then the TEST, TASK and CHOICE columns were recoded into coherent numbers (e.g., TEST 69 included three of the six tasks, TASK 216 included one of two sequences for the alternatives and CHOICE 452 actually meant Lofa as described above). The ACTION column was separated into a column including the information looked at (the INFORMATION-ID) and the keywords searched for.

A.7 Additional figures and tables

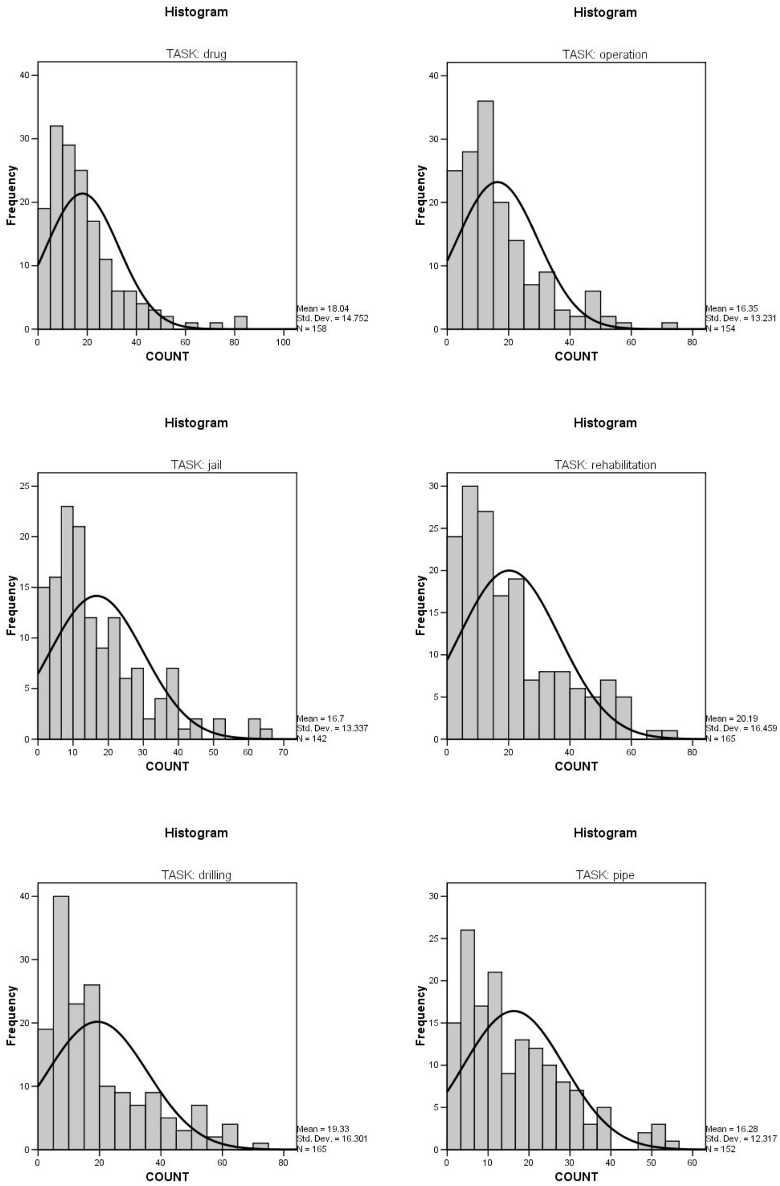


Fig. 35. Frequency distribution of clicks between all tasks

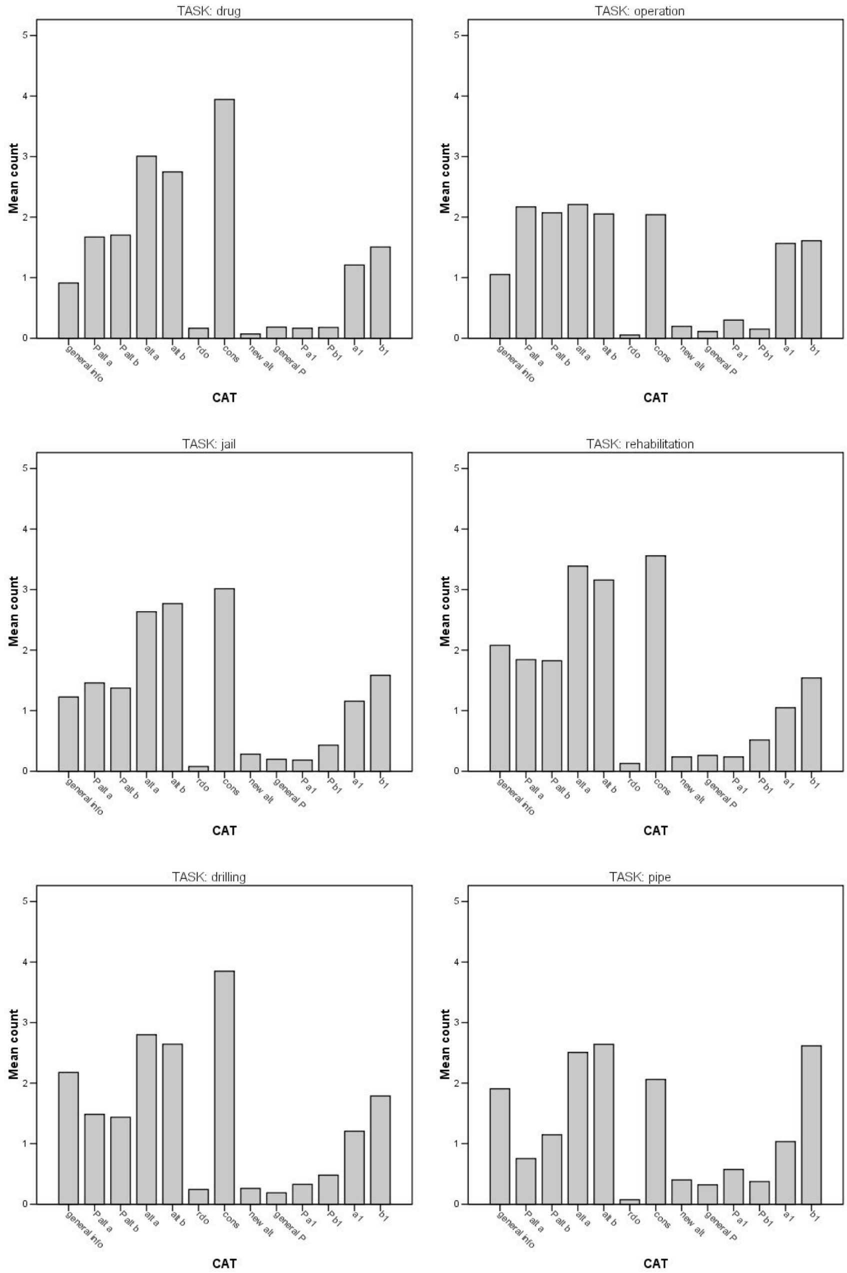


Fig. 36. Mean of categories per task

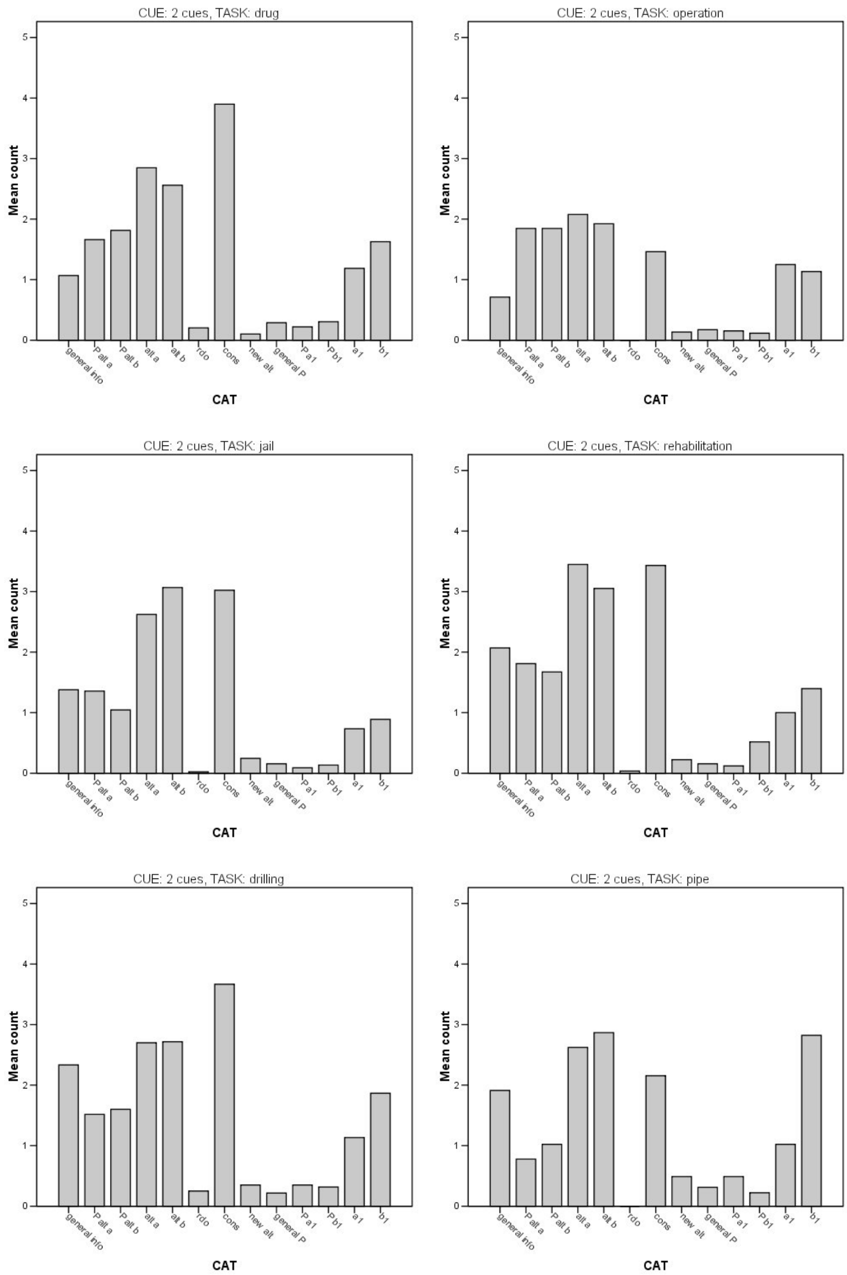


Fig. 37. Condition: 2 cues, all categories

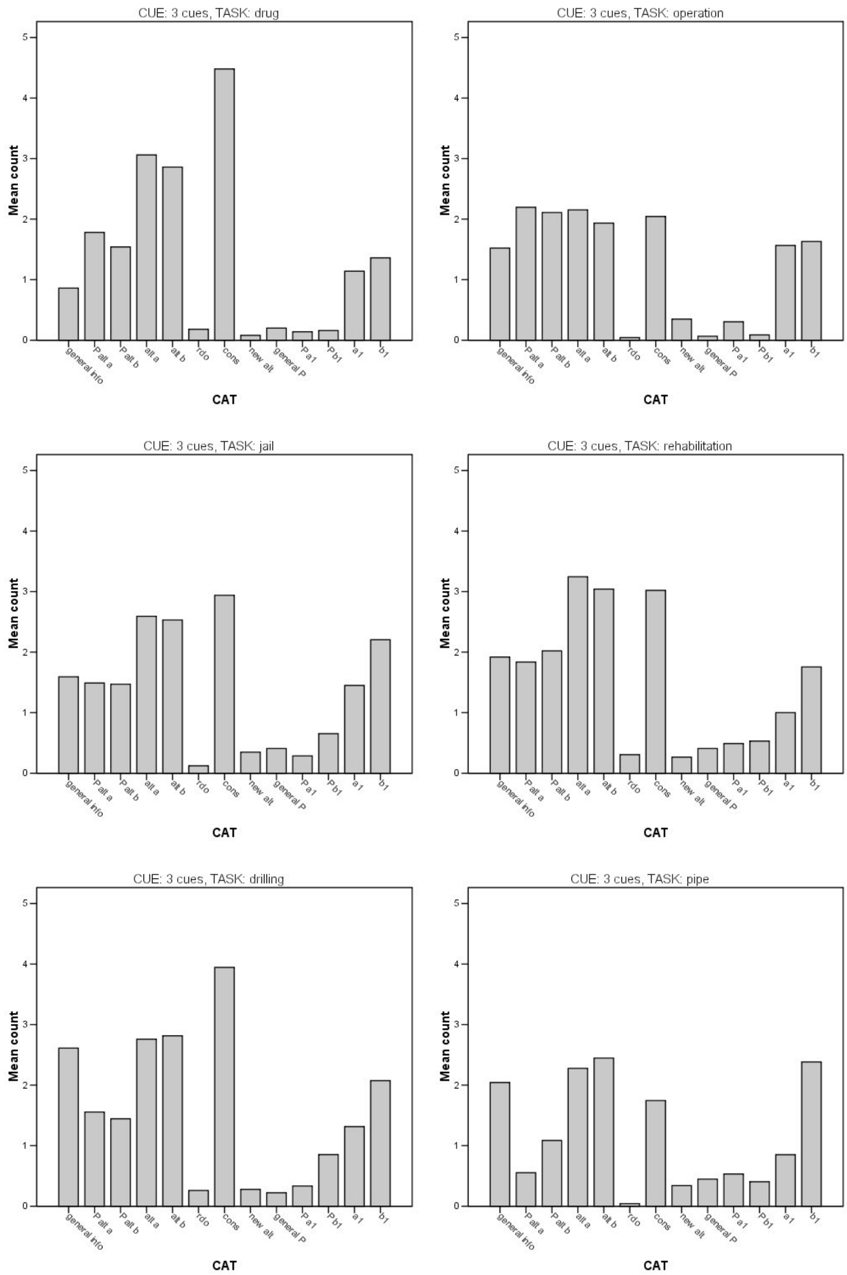


Fig. 38. Condition: 3 cues, all categories

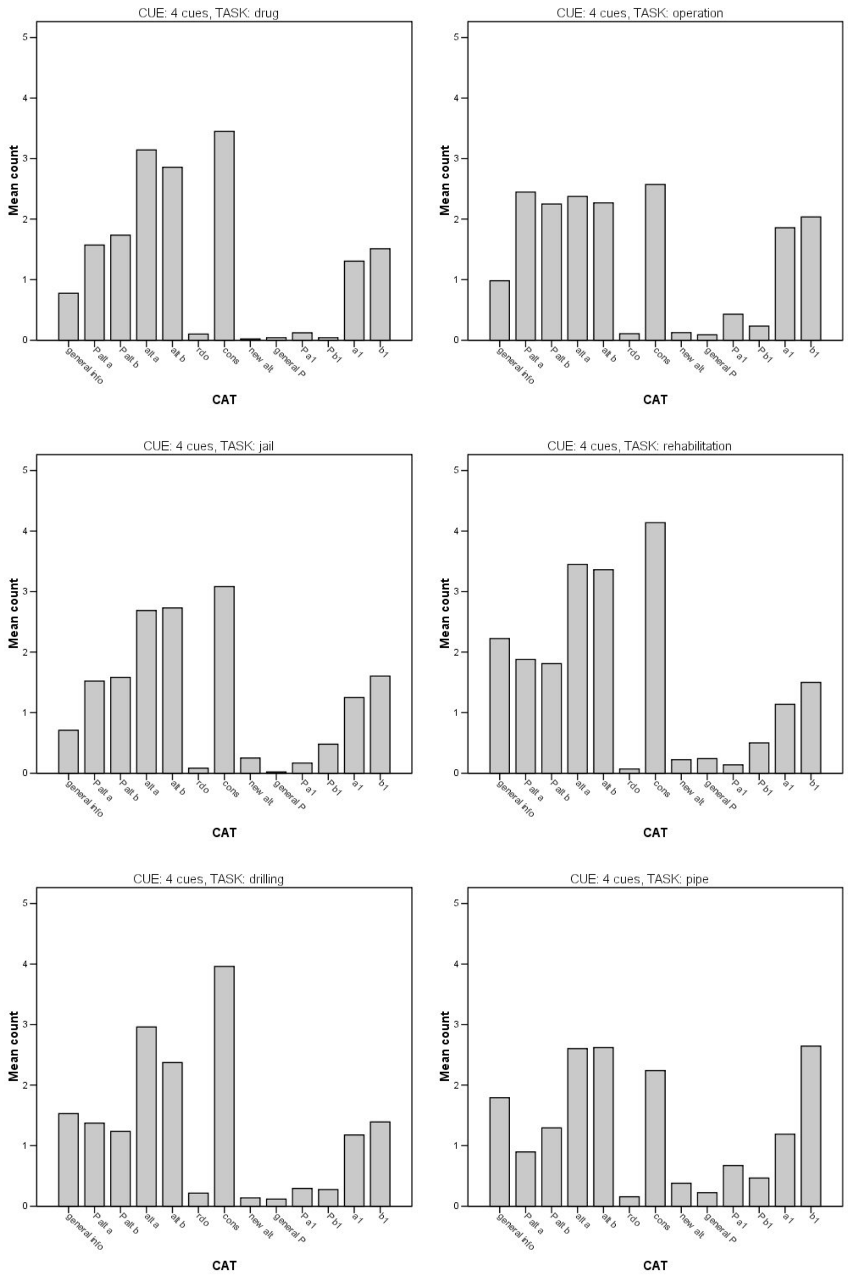


Fig. 39. Condition: 4 cues, all categories

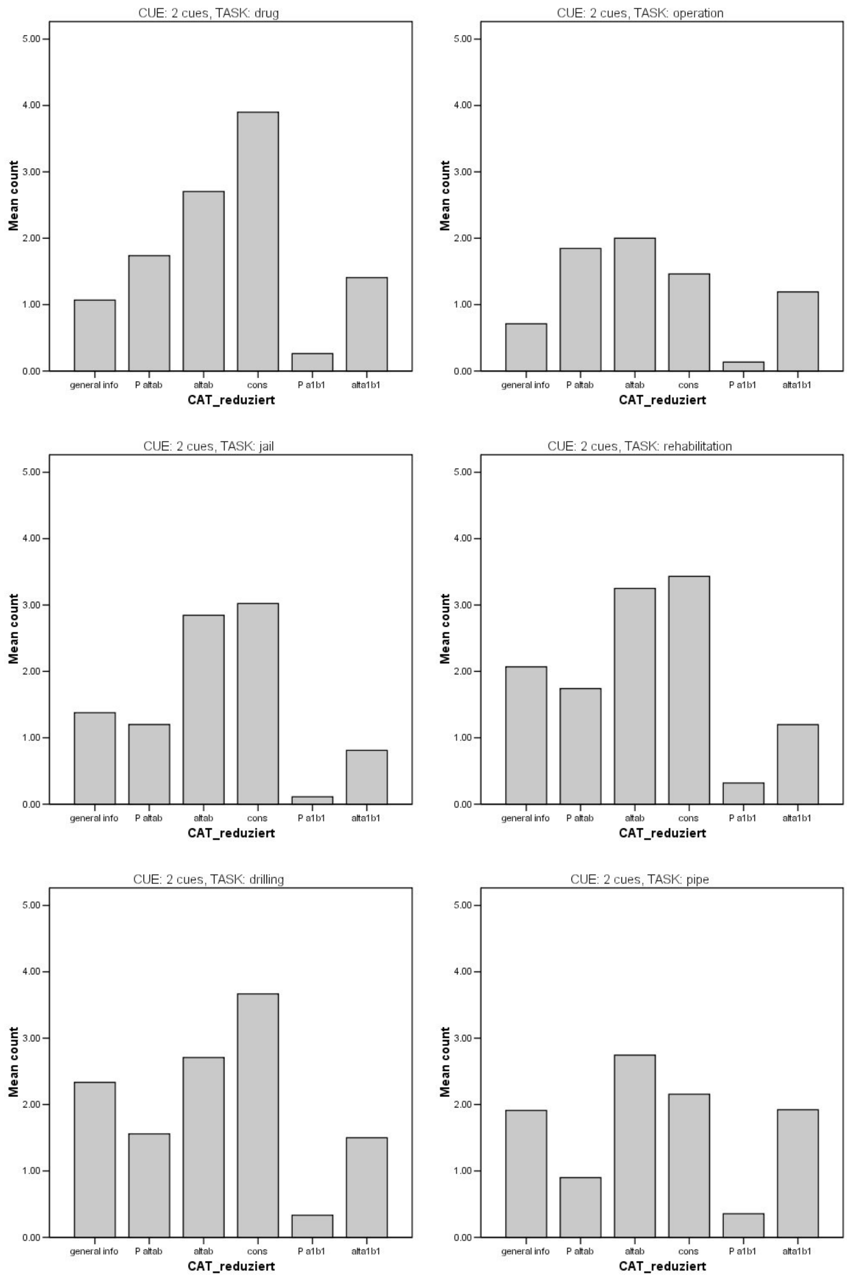


Fig. 40. Condition: 2 cues, selected categories

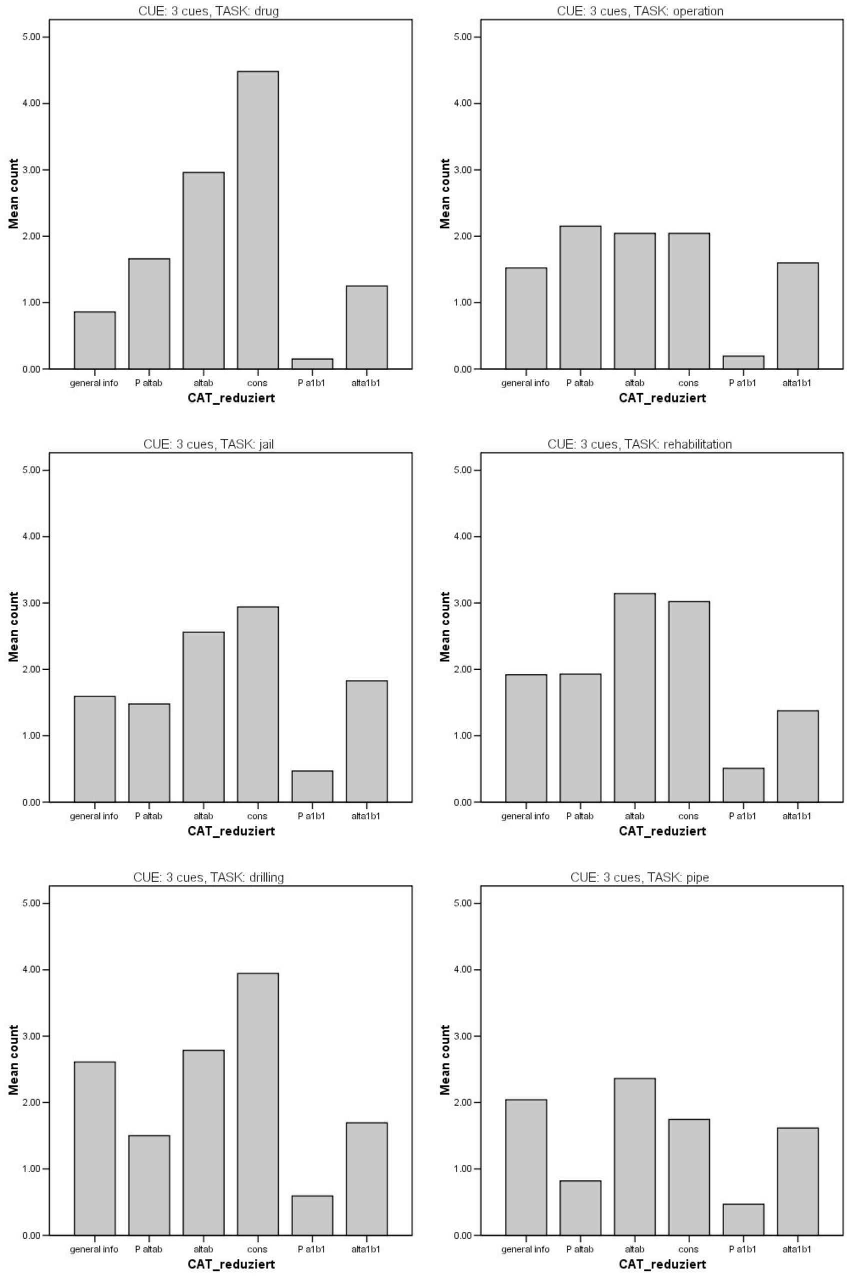


Fig. 41. Condition: 3 cues, selected categories

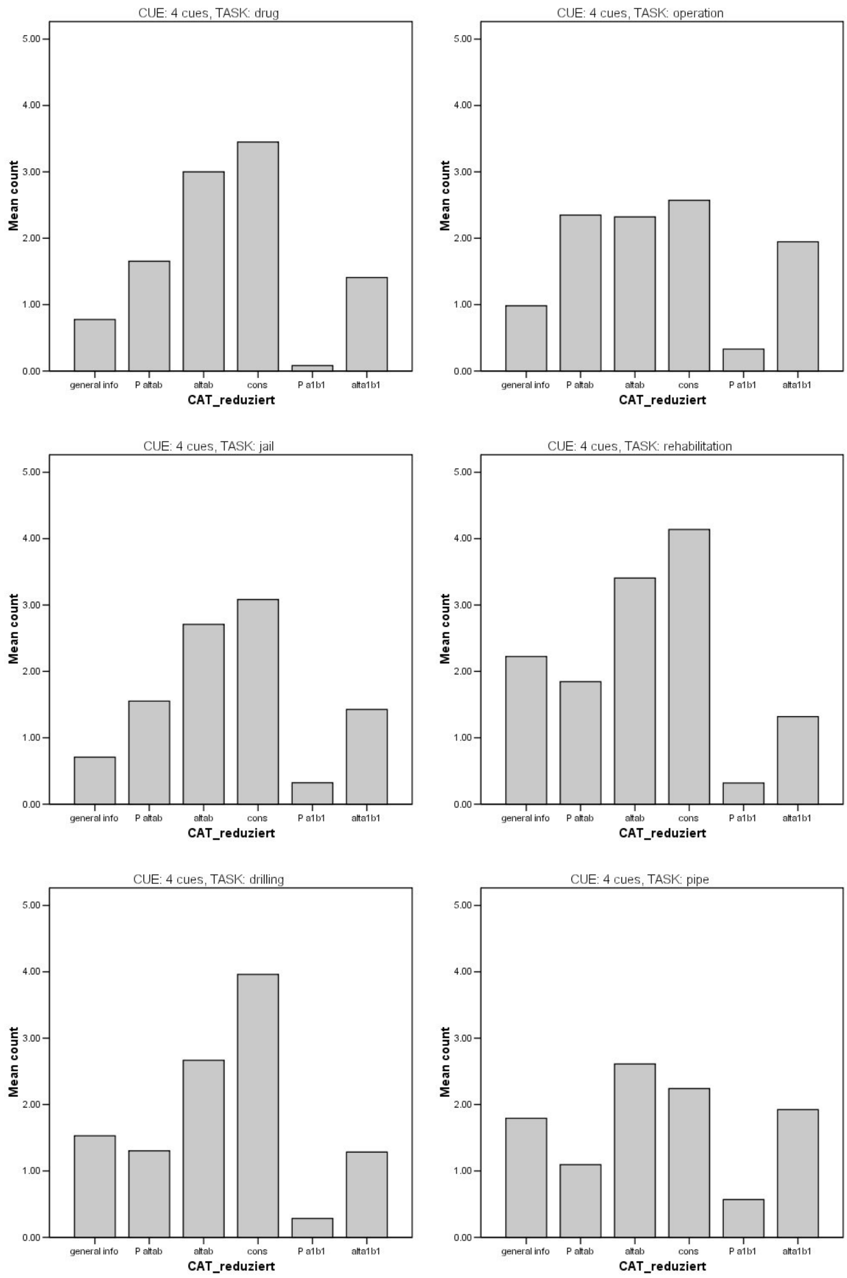


Fig. 42. Condition: 4 cues, selected categories

Correlations[a]

		drilling	pipe	drug	jail	operation	rehabilitation
drilling	Pearson Correlation	1	-.097	.889**	.777**	.739**	.616*
	Sig. (2-tailed)		.753	.000	.002	.004	.025
	N	13	13	13	13	13	13
pipe	Pearson Correlation	-.097	1	-.156	-.015	-.490	-.051
	Sig. (2-tailed)	.753		.610	.961	.089	.869
	N	13	13	13	13	13	13
drug	Pearson Correlation	.889**	-.156	1	.729**	.700**	.879**
	Sig. (2-tailed)	.000	.610		.005	.008	.000
	N	13	13	13	13	13	13
jail	Pearson Correlation	.777**	-.015	.729**	1	.292	.590*
	Sig. (2-tailed)	.002	.961	.005		.334	.034
	N	13	13	13	13	13	13
operation	Pearson Correlation	.739**	-.490	.700**	.292	1	.483
	Sig. (2-tailed)	.004	.089	.008	.334		.095
	N	13	13	13	13	13	13
rehabilitation	Pearson Correlation	.616*	-.051	.879**	.590*	.483	1
	Sig. (2-tailed)	.025	.869	.000	.034	.095	
	N	13	13	13	13	13	13

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

a. CUE = 2 cues

Correlations[a]

		drilling	pipe	drug	jail	operation	rehabilitation
drilling	Pearson Correlation	1	.678*	.860**	.875**	.660*	.837**
	Sig. (2-tailed)		.011	.000	.000	.014	.000
	N	13	13	13	13	13	13
pipe	Pearson Correlation	.678*	1	.664*	.635*	.616*	.504
	Sig. (2-tailed)	.011		.013	.020	.025	.079
	N	13	13	13	13	13	13
drug	Pearson Correlation	.860**	.664*	1	.863**	.702**	.801**
	Sig. (2-tailed)	.000	.013		.000	.007	.001
	N	13	13	13	13	13	13
jail	Pearson Correlation	.875**	.635*	.863**	1	.642*	.870**
	Sig. (2-tailed)	.000	.020	.000		.018	.000
	N	13	13	13	13	13	13
operation	Pearson Correlation	.660*	.616*	.702**	.642*	1	.571*
	Sig. (2-tailed)	.014	.025	.007	.018		.042
	N	13	13	13	13	13	13
rehabilitation	Pearson Correlation	.837**	.504	.801**	.870**	.571*	1
	Sig. (2-tailed)	.000	.079	.001	.000	.042	
	N	13	13	13	13	13	13

*. Correlation is significant at the 0.05 level (2-tailed).

**. Correlation is significant at the 0.01 level (2-tailed).

a. CUE = 3 cues

Correlations[a]

		drilling	pipe	drug	jail	operation	rehabilitation
drilling	Pearson Correlation	1	.652*	.755**	.864**	.776**	.880**
	Sig. (2-tailed)		.016	.003	.000	.002	.000
	N	13	13	13	13	13	13
pipe	Pearson Correlation	.652*	1	.767**	.653*	.629*	.779**
	Sig. (2-tailed)	.016		.002	.016	.021	.002
	N	13	13	13	13	13	13
drug	Pearson Correlation	.755**	.767**	1	.503	.443	.796**
	Sig. (2-tailed)	.003	.002		.080	.129	.001
	N	13	13	13	13	13	13
jail	Pearson Correlation	.864**	.653*	.503	1	.876**	.843**
	Sig. (2-tailed)	.000	.016	.080		.000	.000
	N	13	13	13	13	13	13
operation	Pearson Correlation	.776**	.629*	.443	.876**	1	.698**
	Sig. (2-tailed)	.002	.021	.129	.000		.008
	N	13	13	13	13	13	13
rehabilitation	Pearson Correlation	.880**	.779**	.796**	.843**	.698**	1
	Sig. (2-tailed)	.000	.002	.001	.000	.008	
	N	13	13	13	13	13	13

*. Correlation is significant at the 0.05 level (2-tailed).

**. Correlation is significant at the 0.01 level (2-tailed).

a. CUE = 4 cues

Fig. 43. Correlations of mean clicks per category

Correlations[a]

		drug	jail	drilling	operation	rehabilitation	pipe
drug	Pearson Correlation	1	.934**	.903*	.710	.910*	.672
	Sig. (2-tailed)		.006	.014	.114	.012	.143
	N	6	6	6	6	6	6
jail	Pearson Correlation	.934**	1	.943**	.693	.991**	.802
	Sig. (2-tailed)	.006		.005	.127	.000	.055
	N	6	6	6	6	6	6
drilling	Pearson Correlation	.903*	.943**	1	.574	.961**	.806
	Sig. (2-tailed)	.014	.005		.234	.002	.053
	N	6	6	6	6	6	6
operation	Pearson Correlation	.710	.693	.574	1	.703	.572
	Sig. (2-tailed)	.114	.127	.234		.119	.235
	N	6	6	6	6	6	6
rehabilitation	Pearson Correlation	.910*	.991**	.961**	.703	1	.814*
	Sig. (2-tailed)	.012	.000	.002	.119		.049
	N	6	6	6	6	6	6
pipe	Pearson Correlation	.672	.802	.806	.572	.814*	1
	Sig. (2-tailed)	.143	.055	.053	.235	.049	
	N	6	6	6	6	6	6

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

a. CUE = 2 cues reduced

Correlations[a]

		drug	jail	drilling	operation	rehabilitation	pipe
drug	Pearson Correlation	1	.926**	.866*	.698	.887*	.502
	Sig. (2-tailed)		.008	.026	.123	.019	.310
	N	6	6	6	6	6	6
jail	Pearson Correlation	.926**	1	.919**	.801	.931**	.764
	Sig. (2-tailed)	.008		.009	.056	.007	.077
	N	6	6	6	6	6	6
drilling	Pearson Correlation	.866*	.919**	1	.669	.885*	.758
	Sig. (2-tailed)	.026	.009		.146	.019	.080
	N	6	6	6	6	6	6
operation	Pearson Correlation	.698	.801	.669	1	.828*	.573
	Sig. (2-tailed)	.123	.056	.146		.042	.235
	N	6	6	6	6	6	6
rehabilitation	Pearson Correlation	.887*	.931**	.885*	.828*	1	.747
	Sig. (2-tailed)	.019	.007	.019	.042		.088
	N	6	6	6	6	6	6
pipe	Pearson Correlation	.502	.764	.758	.573	.747	1
	Sig. (2-tailed)	.310	.077	.080	.235	.088	
	N	6	6	6	6	6	6

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

a. CUE = 3 cues reduced

Correlations[a]

		drug	jail	drilling	operation	rehabilitation	pipe
drug	Pearson Correlation	1	.998**	.937**	.885*	.922**	.792
	Sig. (2-tailed)		.000	.006	.019	.009	.060
	N	6	6	6	6	6	6
jail	Pearson Correlation	.998**	1	.923**	.884*	.897*	.776
	Sig. (2-tailed)	.000		.009	.020	.015	.070
	N	6	6	6	6	6	6
drilling	Pearson Correlation	.937**	.923**	1	.734	.973**	.791
	Sig. (2-tailed)	.006	.009		.096	.001	.061
	N	6	6	6	6	6	6
operation	Pearson Correlation	.885*	.884*	.734	1	.732	.648
	Sig. (2-tailed)	.019	.020	.096		.098	.164
	N	6	6	6	6	6	6
rehabilitation	Pearson Correlation	.922**	.897*	.973**	.732	1	.822*
	Sig. (2-tailed)	.009	.015	.001	.098		.045
	N	6	6	6	6	6	6
pipe	Pearson Correlation	.792	.776	.791	.648	.822*	1
	Sig. (2-tailed)	.060	.070	.061	.164	.045	
	N	6	6	6	6	6	6

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

a. CUE = 4 cues reduced

Fig. 44. Correlations of mean clicks per category (reduced)

Table 38. Patterns for clicks - MEDICINE

1	2	%	1	2	3	4	%	1	2	3	4	5	%	1	2	3	4	%
DRUG			OP															
500		71.76	406				66.93	203	406	506			36.22	203	400	406		30.71
400		70.23	506				64.57	201	301	400	406		36.22	201	400	406	500	30.71
500	400	44.27	406	506			62.99	400	406	500	506		35.43	303	400	406	506	29.92
406		39.69	201				60.63	400	500	506			35.43	303	400	506		29.92
506		39.69	400				57.48	203	400				35.43	301	303	506		29.92
200		37.40	301				56.69	201	301	400	406	506	35.43	301	406	500	506	29.92
300		36.64	203				52.76	201	301	400	506		35.43	301	700			29.92
700		35.88	201	301			52.76	201	700				35.43	203	303	400		29.92
301		35.11	303				51.18	1305					34.65	201	203	400		29.92
701		35.11	500				49.61	301	303				34.65	201	301	303		29.92
201		34.35	201	406			49.61	301	500				34.65	201	405	406		29.92
505		33.59	400	406			48.03	201	203	301			34.65					
711		31.30	301	406			47.24	701					33.86					
1305		31.30	201	506			47.24	405	406				33.86					
400	500	30.53	301	506			46.46	303	400				33.86					
405		30.53	201	406	506		46.46	203	303	406			33.86					
714		30.53	400	406	506		45.67	203	303	506			33.86					
1200		30.53	400	506			45.67	201	301	303			33.86					
500	506	29.01	301	406	506		45.67	201	406	500			33.86					
400	406	27.48	201	400			44.88	203	500				33.07					
500	1305	27.48	203	303			44.09	201	303	406			33.07					
712		27.48	201	301	406		44.09	201	405				33.07					
404		26.72	201	301	506		43.31	505	506				32.28					
709		26.72	405				42.52	406	505				32.28					
202		25.95	201	301	406	506	42.52	303	500				32.28					
302		25.95	406	500			41.73	203	301	303			32.28					
400	1200	25.95	400	500			41.73	203	301	406			32.28					
400	405	25.19	301	400			40.94	203	303	406	506		32.28					
402		25.19	505				40.16	201	203	303			32.28					
303		24.43	500	506			40.16	201	301	500			32.28					
400	200	24.43	303	406			40.16	201	303	506			32.28					
500	505	23.66	303	506			40.16	201	400	500			32.28					
400	700	22.90	201	400	406		40.16	201	500	506			32.28					
506	406	22.90	406	500	506		39.37	1202					31.50					
203		22.14	203	406			39.37	406	505	506			31.50					
400	201	22.14	201	301	400		39.37	406	700				31.50					
401		22.14	700				38.58	405	506				31.50					
500	300	22.14	303	406	506		38.58	301	406	500			31.50					
500	301	22.14	301	400	406		37.80	203	301	506			31.50					
500	701	22.14	203	301			37.80	201	203	301	303		31.50					
707		22.14	203	506			37.80	201	203	406			31.50					
708		22.14	201	203			37.80	201	303	406	506		31.50					
500	201	21.37	201	303			37.80	201	406	500	506		31.50					
500	406	21.37	201	400	406	506	37.80	506	700				30.71					
500	707	21.37	201	400	506		37.80	405	406	506			30.71					
503		21.37	201	500			37.80	405	505				30.71					
1201		21.37	400	406	500		37.01	303	400	406			30.71					

Table 39. Patterns for clicks - LAW

	1	2	%		1	2	%	1	2	%
JAIL				REHAB				cont.	task.	REHAB
	400		58.68		400		65.89	301	506	23.26
	500		55.37		500		60.47	400	506	23.26
	406		50.41		701		55.04	506	406	23.26
	506		47.93		406		52.71	303	203	22.48
	700		40.50		506		50.39	700	701	22.48
	701		38.84		700		47.29	400	303	21.71
	1200		35.54		301		45.74	400	502	21.71
	400	406	34.71		203		41.09	400	1200	21.71
	202		33.06		110		38.76	500	110	21.71
	302		33.06		201		37.21	500	700	21.71
	501		30.58		303		35.66	506	400	21.71
	712		30.58		400	203	35.66	709		21.71
	1305		30.58		400	406	35.66	110	406	20.93
	505		29.75		500	506	34.11	201	406	20.93
	405		28.93		405		32.56	301	203	20.93
	500	400	28.93		500	701	31.78	300		20.16
	500	506	28.10		501		31.78	301	201	20.16
	502		28.10		400	500	31.01	301	300	20.16
	709		27.27		1200		30.23	301	406	20.16
	200		26.45		400	700	29.46	500	501	20.16
	300		24.79		402		29.46			
	400	405	23.97		500	301	29.46			
	400	700	23.97		502		29.46			
	401		23.97		1305		29.46			
	406	506	23.97		400	201	28.68			
	500	501	23.97		500	303	27.13			
	400	709	23.14		500	400	27.13			
	400	1200	22.31		500	406	27.13			
	500	302	22.31		505		27.13			
	500	505	22.31		400	405	26.36			
	500	701	22.31		500	203	26.36			
	504		22.31		400	110	25.58			
	201		21.49		400	701	25.58			
	400	506	21.49		406	506	25.58			
	400	200	20.66		400	301	24.81			
	400	202	20.66		400	402	24.81			
	402		20.66		401		24.03			
	500	406	20.66		712		24.03			

Table 40. Patterns for clicks - BUSINESS

	1	2	%		1	2	%
DRILL				PIPE			
	406		64.44		506		61.29
	506		63.70		400		58.87
	701		50.37		500		58.06
	700		49.63		1309		55.65
	500		48.15		120		52.42
	400		48.15		406		51.61
	406	506	37.04		400	500	33.06
	300		36.30		506	120	33.06
	400	406	32.59		1308		33.06
	200		32.59		1201		32.26
	202		30.37		400	1309	29.84
	402		27.41		506	1309	29.84
	1200		26.67		700		29.84
	302		26.67		406	1309	29.03
	709		25.93		500	406	28.23
	702		25.93		500	1308	27.42
	712		24.44		400	506	26.61
	700	701	24.44		500	1309	26.61
	506	406	24.44		701		26.61
	401		24.44		500	1201	25.00
	301		24.44		506	406	25.00
	1305		23.70		703		25.00
	400	700	23.70		400	700	24.19
	500	300	22.96		406	120	23.39
	500	506	22.96		500	120	23.39
	1309		22.22		500	400	23.39
	506	400	22.22		1001		23.39
	502		22.22		120	1309	22.58
	405		22.22		400	406	22.58
	400	500	22.22		400	701	21.77
	1204		21.48		500	506	21.77
	701	506	21.48		303		20.97
	500	701	21.48		406	506	20.97
	400	1200	21.48		502		20.97
	406	500	20.74		126		20.16
	400	202	20.74		201		20.16
	400	506	20.74		302		20.16
	201		20.74		1301		20.16
	700	406	20.00				
	406	300	20.00				
	400	200	20.00				

References

Anderhub, V., Müller, R., & Schmidt, C. (2001). Design and evaluation of an economic experiment via the Internet. *Journal of Economic Behavior and Organization*, *46*, 227-247.

Arroba, T. (1977). Styles of decision making and their use: an empirical study. *British Journal of Guidance and Counceling*, *5*, 149-158.

Ayres, J., Flannick, J., Gehrke, J., & Yiu, T. (2002). Sequential PAttern Mining using a bitmap representation. In *Proceedings of the eighth ACM SIGKDD international conference on knowledge discovery and data mining* (p. 429-435). New York: ACM Press.

Backlund, L., Skaner, Y., Montgomery, H., Bring, J., & Strender, L.-E. (2003). Doctors decision processes in a drug-prescription task: The validity of rating scales and think-aloud reports. *Organizational Behavior and Human Decision Processes*, *91*, 108-117.

Baddeley, A. (1990). *Human memory: Theory and practice.* Hove: Lawrence Erlbaum Associates.

Baddeley, A. (2003). Working memory and language: An overview. *Journal of Communication Disorder*, *36*, 198-208.

Bailey, K. D. (1994). *Typologies and Taxonomies: An introduction to classification techniques.* Thousand Oaks: Sage.

Batinic, B., Reips, U.-D., & Bosnjak, M. (Eds.). (2002). *Online social sciences.* Bern: Hogrefe & Huber.

Birnbaum, M. H. (1999). Testing critical properties of decision making on the Internet. *Psychological Science*, *10*(5), 399-407.

Birnbaum, M. H. (2000). *Psychological experiments on the Internet.* San Diego: Academic Press.

Birnbaum, M. H. (2004). Human research and data collection via the Internet. *Annual Review of Psychology*, *55*, 803-832.

Blais, A.-R., & Weber, E. U. (2001). Domain-specificity and gender differences in decision making. *Risk Decision and Policy*, *6*(1), 47-69.

Browne, G. J., & Pitts, M. G. (2004). Stopping rule use during information search in design problems. *Organizational Behavior and Human Decision Processes*, *95*(2), 208–224.

Buchanan, T., & Smith, J. (1999). Using the Internet for psychological research: Personality testing on the World Wide Web. *British Journal of Psychology*, *90*, 125-144.

Bucklin, R. E., & Sismeiro, C. (2003). A model of Web site browsing behavior estimated on clickstream data. *Journal of Marketing Research*, *15*, 249-267.

Burdick, D., Calimlim, M., & Gehrke, J. (2001). MAFIA: A maximal frequent itemset algorithm for transactional databases. In *Proceedings of the 17th International Conference on Data Engineering.* Heidelberg: CDE.

Cheung, B., & Hofer, K. (2003). Eye tracking, point of gaze, and performance degradation during disorientation. *Aviation, Space, and Environmental Medicine*, *74*(1), 11-20.

Chu, P., & Spires, E. E. (2003). Perceptions of accuracy and effort of decision strategies. *Organizational Behavior and Human Decision Processes*, *91*(2), 203-214.

Dong, J., Martin, S., & Waldo, P. (1999). *A user input and analysis tool for information architecture* (Tech. Rep.).

Engländer, T., & Tyszka, T. (1980). Information seeking in open decision situations. *Acta Psychologica*, *45*, 169-176.

Epstein, J., Klinkenberg, W., Wiley, D., & McKinley, L. (2001). Insuring sample equivalence across Internet and paper-and-pencil assessments. *Computers in Human Behavior*, *17*, 339-346.

Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data.* Cambridge: MIT Press.

Fasolo, B., McClelland, G. H., & Lange, K. A. (in press). The effect of site design and interattribute correlations on interactive Web-based decisions. In C. P. Haugtvedt, K. Machleit, & R. Yalch (Eds.), *Online consumer Psychology: Understanding and influencing behavior in the virtual world.* Mahwah, NJ: Erlbaum.

Fasolo, B., Misuraca, R., & McClelland, G. H. (2003). Individual differences in adaptive choice strategies. *Research in Economics*, *57*(3), 219-233.

Feldman-Stewart, D., Brundage, M. D., Manen, L. V., & Svenson, O. (2004). Patient-focussed decision-making in early-stage prostate cancer: insights from a cognitively based decision aid. *Health Expectations*, *7*, 126-141.

Fisseni, H.-J. (2004). *Lehrbuch der psychologischen Diagnostik* (3. ed.). Göttingen: Verlag für Psychologie.

Ford, J., Schmitt, N., Schechtmann, S., Hults, B., & Doherty, M. (1989). Process tracing methods: Contributions, problems, and neglected research questions. *Organizational Behavior and Human Decision Processes*, *43*, 75-117.

Frederiksen, N. (1972). Toward a taxonomy of situations. *American Psychologist*, *2*, 114-123.

Freedman, D. (1983). *Markov chains.* New York: Springer.

Frick, A., Bächtiger, M.-T., & Reips, U.-D. (2001). Financial incentives, personal information, and drop out in online studies. In U.-D. Reips & M. Bosnjak (Eds.), *Dimensions of Internet science* (p. 209-219). Lengerich: Papbst Science Publishers.

Gangestad, S., & Snyder, M. (1985). To carve nature at its joints: On the existence of discrete classes in personality. *Psychological Review*, 317-340.

Gigerenzer, G., & Selten, R. (2001). *Bounded rationality: The adaptive toolbox.* Cambridge: The MIT Press.

Gigerenzer, G., Todd, P. M., & ABC Research Group the. (1999). *Simple heuristics that make us smart.* Oxford: Oxford University Press.

Gilgen, K., & Schwizer, R. (2003). *Taxonomien individueller Entscheidungen: Hinweise zu aktueller Literatur zwischen 1993 und 2003.* (Seminararbeit: Universität Fribourg)

Goldstein, W., & Weber, E. (1995). Content and discontent: Indications and implications of domain specificity in preferential decision making. In J. Busemeyer, D. Medin, & R. Hastie (Eds.), *Decision making from a cognitive perspective* (Vol. 32, p. 83-136). Academic Press.

Gosling, S. D., Vazire, S., Srivastava, S., & John, O. P. (2004). Should we trust Web-based studies? A comparative analysis of six preconceptions about Internet questionnaires. *American Psychologist*, *59*(2), 93-104.

Harte, J. M., & Koele, P. (2001). Modelling and describing human judgment processes: The multiattribute evaluation case. *Thinking and Reasoning*, *7*(1), 29-49.

Harte, J. M., Westenberg, M. R., & Someren, M. van. (1994). Process models of decision making. *Acta Psychologica*, *87*, 95-120.

Huber, O. (1979). Nontransitive multidimensional preferences: Theoretical analysis of a model. *Theory and Decision*, *10*, 147-165.

Huber, O. (1997). Beyond gambles and lotteries: Naturalistic risky decisions. In R. Ranyard, W. Crozier, & O. Svenson (Eds.), *Decision making: Cognitive models and explanations* (p. 145-162). London: Routledge.

Huber, O. (2004). Risky decision making: focus on risk defusing behavior. In F. Columbus (Ed.), *Advances in Psychology research.* New York: Nova Science Publishers.

Huber, O., Beutter, C., Montoya, J., & Huber, O. W. (2001). Risk defusing behaviour: Towards an understanding of risky decision making. *European Journal of Cognitive Psychology*, *13*(3), 409-426.

Huber, O., & Huber, O. W. (submitted). What makes gambles special? - Active search for probability and risk defusing information in gambles and quasi-realistic risky tasks. *Organizational Behavior and Human Decision Processes.*

Huber, O., & Macho, S. (2001). Probabilistic set-up and the search for probability information in quasi-naturalistic decision tasks. *Risk Decision and Policy*, *6*, 1-16.

Huber, O., Wider, R., & Huber, O. W. (1997). Active information search and complete information presentation in naturalistic risky decision tasks. *Acta Psychologica*, *95*, 15-29.

Jasper, J., & Shapiro, J. (2002). MouseTrace: A better mousetrap for catching decision processes. *Behavior Research Methods, Instruments & Computers*, *34*(3), 364-374.

Johnson, E. J. (2001). Digitizing consumer research. *Journal of Consumer Research*, *28*, 331-336.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decisions under risk. *Econometrica*, *47*, 263-291.

Klein, G. (1993). A recognition-primed decision (RPD) model of rapid decision making. In G. Klein, J. Orasanu, R. Calderwood, & C. E. Zsambok (Eds.), *Decision making in action: Models and methods* (p. 138-147). Norwood: Ablex Publishing Corporation.

Klein, G., Orasanu, J., Calderwood, R., & Zsambok, C. (Eds.). (1995). *Decision making in action: Models and methods.* New Jersey: Norwood.

Krantz, J., Ballard, J., & Scher, J. (1997). Comparing the results of laboratory and World-Wide Web samples on the determinants of female attractiveness. *Behavior Research Methods, Instruments, & Computers*, *29*(2), 264-269.

Krantz, J., & Dalal, R. (2000). Validity of Web-based psychological research. In M. Birnbaum (Ed.), *Psychological experiments on the Internet* (p. 35-60). San Diego: Academic Press.

Kühberger, A. (1998). The influence of framing on risky decision making: A meta-analysis. *Organizational Behavior and Human Decision Processes*, *75*(1), 23-55.

Kühberger, A., Schulte-Mecklenbeck, M., & Perner, J. (2002). Framing decisions: hypothetical and real. *Organizational Behavior and Human Decision Processes*, *89*, 1162-1175.

Levin, I. P., Huneke, M., & Jasper, J. (2000). Information processing at successive stages of decision making: Need for cognition and inclusion-exclusion effects. *Organizational Behavior and Human Decision Processes*, *82*(2), 171-193.

Levin, I. P., Schneider, S., & Gaeth, G. (1998). All frames are not created equal: A typology and critical analysis of framing effects. *Organizational Behavior and Human Decision Processes*, *76*, 149-188.

Loeber, S., & Cristea, A. (2003). A WWW information seeking process model. *Educational Technology & Society*, *6*(3), 43-52.

Loewenstein, G., Weber, E., Hsee, C., & Welch, N. (2001). Risk as feelings. *Psychological Bulletin*, *127*(2), 267-286.

Lohse, G., & Johnson, E. J. (1996). A comparison of two process tracing methods for choice tasks. *Organizational Behavior and Human Decicion Processes*, *68*(1), 28-43.

Lüer, G., Lass, U., & Shallo-Hoffmann, J. (1988). *Eye movement research.* Toronto: Hogrefe.

Maule, A. J. (1989). Positive and negative decision frames: A verbal protocol analysis of the Asian disease problem of Tversky and Kahneman. In H. Montgomery & O. Svenson (Eds.), *Process* *and structure in human decision making* (p. 163-180). Chichester: John Wiley & Sons.

Maule, A. J., & Pearman, A. (1994). *Taxonomies of decision making* (Tech. Rep.). Centre for Decision Research Leeds University Business School.

McGraw, K., Tew, M., & Williams, J. (2000). The integrity of Web-delivered experiments: Can you trust the data? *Psychological Science*, *11*(6), 502-506.

Medin, D. L., Schwartz, H. C., Blok, S. V., & Birnbaum, L. A. (1999). The semantic side of decision making. *Psychonomic Bulletin & Review*, *6*(4), 562-569.

Mehta, R., & Sivada, E. (1995). Comparing response rates and response content in mail vs. electronic mail surveys. *Journal of the Market Research Society*, *17*(4), 429-440.

Montgomery, A. L., Li, S., Srinivasan, K., & Liechty, J. C. (in press). Modeling online browsing and path analysis using clickstream data. *Marketing Science*.

Montgomery, H. (1989). From cognition to action: The search for dominance in decision making. In H. Montgomery & O. Svenson (Eds.), *Process and structure in human decision making* (p. 23-51). Chichester: John Wiley & Sons.

Montgomery, H., & Svenson, O. (1989a). *Process and structure in human decision making*. Chichester: John Wiley & Sons.

Montgomery, H., & Svenson, O. (1989b). A think-aloud study of dominance structuring in decision processes. In *Process and structure in human decision making* (p. 135-150). Chichester: John Wiley & Sons.

Montgomery, H., & Willen, H. (1999). Decision making and action: The search for a good structure. In P. Juslin & H. Montgomery (Eds.), *Judgment and decision making: Neo-Brunswikian and process-tracing approaches* (p. 147-173). Mahwah: Lawrence Erlbaum.

Musch, J., & Klauer, K. (2002). Psychological experiments on the World Wide Web: Investigating content effects in syllogistic reasoning. In B. Batinic, U.-D. Reips, & M. Bosnjak (Eds.), *Online social sciences*. Göttingen: Hogrefe.

Musch, J., & Reips, U. D. (2000). A brief history of Web experimenting. In M. H. Birnbaum (Ed.), *Psychological experiments on the internet* (p. 61-85). San Diego: Academic Press.

Neun, M. (2004). *Web based decsion process tracing and mining*. Unpublished master's thesis, University of Fribourg.

Newell, B. R., Weston, N. J., & Shanks, D. R. (2003). Empirical tests of the fast-and-frugal heuristic: Not everyone takes-the-best. *Organizational Behavior and Human Decision Processes*, *91*, 82-96.

Nisbett, R., & Wilson, T. (1977). Telling more than we can know: verbal reports on mental processes. *Psychological Review*, *84*, 231-259.

Osborne, J. W., & Overbay, A. (2004). The power of outliers (and why researchers should always check for them). *Practical Assessment, Research & Evaluation*, *9*(6), 1-12.

Payne, J. W. (1976). Task complexity and contingent processing in decision making: An information search and protocol analysis. *Organizational Behavior and Human Decicion Processes*, *16*, 366-387.

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1993). *The adaptive decision maker*. New York: Cambridge University Press.

Payne, J. W., & Braunstein, M. L. (1978). Risky choice: An examination of information aquisition behavior. *Memory & Cognition*, *6*(5), 554-561.

Pers, J., Bon, M., Kovacic, S., Sibila, M., & Dezman, B. (2002). Observation and analysis of large-scale human motion. *Human Movement Science*, *21*(2), 295-311.

Plous, S. (1993). *The Psychology of judgment and decision making*. New York: McGraw Hill.

Pohl, R., Bender, M., & Lachmann, G. (2002). Hinsight bias around the world. *Experimental Psychology*, *49*, 270-282.

Reips, U.-D. (2000). The Web experiment method: Advantages, disadvantages, and solutions. In M. Birnbaum (Ed.), *Psychological experiments on the Internet* (p. 89-114). San Diego: Academic Press. (very good article)

Reips, U.-D. (2002). Internet-based psychological experimenting. *Social Science Computer Review*, *20*(3), 241-249.

Reips, U.-D., & Bosnjak, M. (2001). *Dimensions of Internet science*. Lengerich: Pabst Science Publishers.

Rettinger, D., & Hastie, R. (2001). Content effects on decision making. *Organizational Behavior and Human Decicion Processes*, *85*(2), 336-359.

Rieskamp, J., & Hoffrage, U. (1999). When do people use simple heuristics, and how can we tell? In G. Gigerenzer, P. M. Todd, & the ABC Research Group (Eds.), *Simple heuristics that make us smart* (p. 141-167). Oxford: Oxford University Press.

Saad, G., & Russo, J. E. (1996). Stopping criteria in sequential choice. *Organizational Behavior and Human Decision Processes*, *67*(3), 258–270.

Schulte-Mecklenbeck, M. (2004). Brave new World ... Wide Web. Blending old teaching methods with a cutting-edge virtual learning environment. *APS Observer*, *17*(10), 48-53.

Schulte-Mecklenbeck, M., & Huber, O. (2003). With or without the experimenter: Information search in the laboratory and on the Web. *Behavior Research Methods, Instruments, & Computers*, *35*(2), 227-235.

Schulte-Mecklenbeck, M., & Neun, M. (in press). WebDiP - A tool for tracing decision making experiments on the WWW. *Behavior Research Methods*.

Shavit, T., Sonsino, D., & Benzion, U. (2001). A comparative study of lotteries-evaluation in class and on the web. *Journal of Economic Psychology*, *22*, 484-491.

Shilo, S., Gerad, L., & Goldman, B. (under review). Patients information needs and decision-making processes: what can be learned from genetic counceling?

Slovic, P. (1987). Perception of risk. *Science*, *236*, 280-285.

Slovic, P. (2001). *The perception of risk*. New York: Earthscan Publications.

Stokmans, M. (1992). Analyzing information search patterns to test the use of a two-phased decision strategy. *Acta Psychologica*, *80*, 213-227.

Svenson, O. (1989). Eliciting and analysing verbal protocols in process studies of judgment and decision making. In H. Montgomery & O. Svenson (Eds.), *Process and structure in human decision making* (p. 65-83). Chichester: John Wiley & Sons.

Svenson, O. (1990). Some propositions for the classification of decisions situations. In K. Borcherding, O. Larichev, & D. Messick (Eds.), *Contemporary issues in decision making* (p. 17-31). Amsterdam: Elsevier.

Svenson, O. (1992). Differentiation and consolidation theory of human decision making: A frame of reference for the study of pre- and post-decision processes. *Acta Psychologica*, *80*, 143-168.

Svenson, O. (1996). Decision making and the search for fundamental psychological regularities: What can we learn from a process perspective. *Organizational Behavior and Human Decision Processes*, *65*(3), 252-267.

Svenson, O. (1999). Differentiation and consolidation theory: Decision making processes before and after a choice. In P. Juslin & H. Montgomery (Eds.), *Judgment and decision making: Neo-Brunswikian and process-tracing approaches* (p. 175-197). Mahwah: Lawrence Erlbaum.

Tada, Y., & Weber, E. U. (1998). Representing psychological dimensions of decisions: Implications for behavioral decision models. In M. A. Gernsbach & S. J. Derry (Eds.), *Proceedings of the twentieth annual conference of the cognitive science society* (p. 1049-1054). Hillsdale, NJ: Lawrence Erlbaum.

Thaler, R. H. (1985). Mental accounting and consumer choice. *Marketing Science*, *4*, 199-214.

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the Psychology of choice. *Science*, *211*, 453-458.

Vaughan, J. (2004). A Web-based archive of norms, stimuli, and data. *Behavior Research Methods, Instruments, & Computers*, *36*(3), 363-370.

Wagenaar, W., Keren, G., & Lichtenstein, S. (1988). Islanders and hostages: Deep and surface structures of decision problems. *Acta Psychologica*, *67*, 175-189.

Weber, E. U., Blais, A., & Betz, N. (2002). A domain-specific risk-attitude scale: Measuring risk perceptions and risk behaviors. *Journal of Behavioral Decision Making*, *15*(4), 263-290.

Weber, E. U., & Hsee, C. (1998). Cross-cultural differences in risk perception, but cros-cultural similarities in attitudes towards perceived risk. *Management Science*, *44*(9), 1205-1217.

White, M., & Iivonen, M. (2001). Questions as a factor in Web search strategy. *Information Processing and Management*, *37*, 721-740.

Willemsen, M. C., & Johnson, E. J. (2004, November). *MouselabWEB: performing sophisticated process tracing experiments in the participant?s home!* (Society for Computers in Psychology (SCIP) annual meeting)

Williamson, J., Ranyard, R., & Cuthbert, L. (2000). Risk management in naturalistic insurance decisions: Evidence from a simulation study. *Risk Decision & Policy*, *5*, 19-38.

Yates, J., & Lee, J. (1996). Chinese decision-making. In M. H. Bond (Ed.), *The handbook of Chinese Psychology* (p. 338-351). Hong Kong: Oxford University Press.

Zsambok, C., & Klein, G. (Eds.). (1997). *Naturalistic decision making*. New Jersey: Erlbaum.