# The paradox of the guided user: assistance can be counter-effective

De paradox van hulp aan de gebruiker: assistentie kan averechts werken
(met een samenvatting in het Nederlands)

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de Universiteit Utrecht op gezag van de rector magnificus, prof.dr. H. Stoof, ingevolge het besluit van het college voor promoties in het openbaar te verdedigen op maandag 31 maart 2008 des middags te 16.15 uur

door

Christof van Nimwegen
geboren op 9 augustus 1971
te Nürnberg, Duitsland

Promotor: Prof. dr. J. van den Berg

Co-promotor: Dr. H. van Oostendorp

SIKS Dissertation Series No. 2008-09
The research reported in this thesis has been carried out under the auspices of SIKS, the Dutch Research School for Information and Knowledge Systems.

Cover design and photography by Laura Chavarria
Printed by Gildeprint, Enschede, The Netherlands

ISBN: 978-90-393-4778-2
Copyright © 2008 Christof van Nimwegen

## Acknowledgements

Let me start with a word about my co-promotor Herre van Oostendorp who was my daily supervisor, and with whom I go back a long way. When I received my Master Degree in 1999, he ended his speech by saying “maybe we’ll work together again someday”. Years after, in 2002, when I returned from living abroad I got a junior teaching job at Utrecht University via Herre. We spoke about doing a PhD. several times that year and when he asked me whether I was interested in this particular project, I did not think for long. Herre, thank you for having the confidence to endeavor this project with me, and for the tremendous effort you invested (also outside working hours) in helping me setup experiments and endlessly rereading my texts. You allowed me enormous freedom, without letting me forget what it is we were aiming at. You let me swim at times, but also, more importantly pushed me the right way when I was under the risk of deviation. In that same vein, in my opinion, the best lesson you constantly taught me, is the importance of realizing that results that might become facts, are worthless without proper interpretation. It was nice to work with you, and now I can also say, who knows, perhaps we’ll work together again!

Two other persons from Utrecht University that were close to the project have to be mentioned. The first one is Hermi Tabachneck-Schijf, who was also at the cradle of this project, and was involved for more than half of the duration. The second one is of course Jörgen van den Berg, who besides giving confidence and being my promoter, was there for me as a “peoples” man when it was necessary. I also thank the reading committee, Peter Gerjets, Paul Kirschner, Mark Neerincx, Arno Siebes and Gerrit van der Veer for their effort.

I especially want to thank Daniel Burgos (then working at the Open University) for the very pleasant and fruitful cooperation and for the coding of the applications, and students Imke de Jong, Marijn Plomp and Jasper Roques for helping me with the experiments.

I had a great time here with my colleagues from our group, of which some became friends. I especially thank Ion Juvina (now at CMU Pittsburgh) for helping me with some statistical analyses and Marco Wiering for helping me with the formulas for the strategy measures, and thank you both for the good conversations (not only concerning science). From Information Science: Emiel (now at TNO), Hans, Eleonore, Joske, Frans, Mari Carmen, Anita, Rik, Virginia, RJ, Rogier and Huub, it was very nice to have you around, and from outside our group, I mention Jeroen, Rainer and Wim.

Of course my gratitude in general goes to the Institute of Information and Computing Sciences that allowed me to do my research in freedom, and facilitated the courses I attended and the scientific arenas I went to present and discuss work with international fellow researchers. I also mention the SIKS office, and the courses I followed organized by Richard Starmans that were very informative and enjoyable.

To the world outside the University walls, I am glad to have had interesting questions, frowns, and more the like from my family and friends who are simply there for me. This brings me to the end of this little prose, and thereby to the most important part, my own family. Laura, thank you for your support in many ways, you are one of those people who is able to support the significant other in justly the right way: sometimes by saying nothing, and sometimes by asking the right questions, and above all, continuously give one the feeling of confidence. In April last year our little girl Yara was born, and I although she did not write one single letter, she has been a positive force for me in finishing up, relatively within time.

# Contents

# 1 Introduction

Computer systems, software and web applications have seen a tremendous development during the last decades, becoming more advanced, and demanding fast learning from users (Neerincx & Griffioen, 1996). Only 40 years ago, computers were huge machines that could only be operated by skilled experts who had to have active command of hundreds of functions. A revolution in workplaces took place in the sense that unskilled people suddenly had (or wanted) to work with computers. Their profession was not primarily geared to the computing medium itself, but the computer became a tool in their everyday work. However, these people were frustrated by the difficulty of learning how to program computers. With the advent of the personal computer during the 1980's it was obvious that their widespread acceptance would become more dependent on their ease of learning and use. Shortly after, when the PC found its way into people's homes, the notion of usability came along (Gould & Lewis, 1983; Shackel, 1991). It was quite an evolution from that huge calculator that hardly fitted in a room, to the laptops and even smaller handheld devices that crowd our environment today. A similar evolution took place in software development. The interfaces of today are by no means comparable with command-line interfaces of the past. Computers are used for countless tasks that had some equivalent tool before (typewriting, chatting, photo/video editing, accounting etc.). Computers abound, also at schools where children use them and play with language or scientific principles and it is hard to imagine any training environment (firemen, pilots, policemen, the military etc.) where computers are *not* used as an aid. Also in large complex systems such as power plant processing systems or airplanes, a great deal of the tasks is aided, or is taken over by computer systems, often in such a way that the user of the system can devote a maximum of attention to other processes.

Many computer-based tasks can in essence (at the very least while learning them) be regarded as problem solving situations, and these days we are used to rather friendly modern systems in which common human computer interaction principles are implemented. In the past, most of the knowledge to work with a computer system had to come from the memory of the user, and users had to be in constant command of an often large body of knowledge. Nowadays, systems are much more intuitive, friendly, and designed in such a way that a broad range of users is able to efficiently use them. In modern computer interfaces, it is common to work in a more *display-based* manner, and *recognition* rather than *recall* of commands and interface objects drives the interaction. Users greatly rely on the interface and the support and guidance that can be derived from it. Examples are systems where users are provided with continuous visual feedback and guidance, and allowing for errors (in the sense that undoing actions is possible) is a common feature in a broad range of software.

However, there are situations where error-recovery is *not* high on the list, but where the exclusion of the chance that users make errors in the first place is important. Can it be fruitful to sometimes make a system that is *not* as assisting and guiding as one is used to? For example, in safety critical systems, one might on the one hand want to exclude the possibility of ambiguous information given by the software, and to safeguard error prevention. On the other hand, while working with these systems, human controllers should at all times have necessary knowledge at their immediate command. The systems should safeguard that controllers are continuously proactive, and do not operate in a shallow manner caused by a system that gives users the feeling the task is being carried out for them. It is exactly these issues that we investigate. The crux of this research is to examine the circumstances under which different interface presentations lead to the best performance in learning a task, performance once the task is learned, and performance under new circumstances, and to the best long-term memory of the task. In addition, the influence of interruptions in the workflow, and instructions on performance will be studied.

## 1.1 Usability research

Since this research takes place in a Human Factors setting it is logical to mention the term *Usability*, which became especially prevalent alongside the introduction of the PC in people's homes. An important shift in paradigm was the introduction of the Graphical User Interface (GUI), a user interface that allows users to interact with a computer that employs graphical icons, visual indicators, and special graphical elements. Using the *desktop metaphor* (Apple, 1987), concepts were represented as objects that match users' physical environment, such as a desktop with items on it, a garbage bin, folders as if they were file cabinets, and many other features. The windows environment in which most people work nowadays is an example of this. Along with the GUI came the interaction paradigm of *direct manipulation*, an interaction style that involves continuous representation of objects of interest, and rapid, reversible, incremental actions and feedback (Shneiderman, 1983). Users are allowed to directly manipulate objects presented to them, performing actions that to some extent correspond to the physical world. Central to this paradigm is that most actions are performed by clicking, drawing, dragging and dropping objects with the mouse. Usability as such is an issue, which has been on the rise in software, websites, and systems development. It originated in the 1990's, and has dramatically changed the way that products are designed and manufactured. Still, usability issues were important already a long time before that. Classic examples of products with usability problems were the ever-famous VCR that nobody managed to program correctly, car dashboards, poorly designed elevator panels, ATM's or kitchen stoves and so on. Cases that are more dramatic were ferry, traffic, airplane, or power plant disasters in which confusing systems lead to human error causing loss of lives. In the common ICT domain, one can think of the ease of use in off-the-shelf software in general, navigation in complex websites, or interactions with mobile devices such as mobile phones, PDA's, and GPS systems. The focus is always on how to make products more intuitive and easy to use. Anyone should be able to buy a subway ticket from a ticket machine, find needed information on a municipality website, or use office software without taking an intensive course first. The systems of today have rather natural language dialogues with us such as "are you sure you want to …." or "the disk you inserted does not have sufficient disk space". This contrasts with MS-DOS-like messages such as "bad or missing command interpreter", "internal stack overflow", or "invalid media, track 0 bad or unusable". In the same vein, when installing an application in MS-Windows or Mac OS, the routine for the user mostly consists of inserting a disc, clicking "yes" or "no" at various instances, checking radio buttons, tagging checkboxes and so on. What is often seen in these interfaces are situations where there are two buttons, e.g. "cancel" and "next". "Next" would be grayed out, indicating that it is impossible to click on at the moment. One would have to check an option (e.g. "I agree"), and only *then,* continuing (by clicking on "next") is possible. Interactions are as smooth as can be, allowing laymen to install applications, without having to worry about deeper system-related issues. Making systems more easy to use in itself is a very good thing, and user satisfaction is nowadays higher than before. Pioneers in the usability field were Jacob Nielsen, Donald Norman, Ben Shneiderman, William Gaver, and many others at research institutes such as Xerox (Palo Alto), MIT and the University of Maryland. A book that had quite an impact was Jacob Nielsen's Usability Engineering (1993), after which the term usability was more firmly established than ever before. Besides the "noble" motivation of product improvement, the broad public realized to some extent that usability goals are business goals, e.g. websites that are hard to use frustrate customers, forfeit revenue, and can be the reason of company bankrupts. Rubin stated in the preface of his book (1994, p. xv):

*"Usability is a business phenomenon. Whole industries and businesses have sprung up to help us operate computer-based products and systems, electronic equipment, and even household appliances that we own, but cannot use properly".*

Most people would agree that the simple idea (simple in intention, not in practice) to improve the products we use in our daily lives, is a positive development. Although definitions of usability are plentiful, we mention the one from ISO standard 9241-11:

*"Usability: the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use".*

Despite the various definitions, broadly at least five characteristics of usability are agreed on:

- **Learnability:** How easy is it for users to accomplish tasks the first time they encounter the design?
- **Efficiency of use:** Once users have learned the design, how quickly can they perform tasks?
- **Memorability:** When users return to the design after a period of not using it, how easily can they re-establish proficiency?
- **Few and non-catastrophic errors:** How many errors do users make, how severe are these errors, and how easily can they recover from the errors?
- **Satisfaction:** How pleasant is it to use the design?

A recurring issue in usability guidelines is the importance of minimizing "user memory load", also referred to as *computational offloading* (Scaife & Rogers, 1996). This means that the working memory (WM) of a user is relieved so that a maximum of cognitive resources can be devoted to the task. To achieve this, certain information can be *externalized* to the problem solver and thus be of assistance. Already two decades ago, Norman (1988) proposed the idea that knowledge might be as much in the world as it is in the head. He pointed out that the information embedded in technological artifacts (such as computer interfaces) was as important to task achievement as the knowledge residing in the mind of the user. Norman argued that well-designed artifacts that externalized information concerning their functions could reduce users' memory load, while badly designed artifacts increased the demands made on the user. Zhang and Norman (2004) and Zhang (1997) characterized external representations as constraining the range of possible cognitive behaviors in the sense that some behaviors are allowed and others prohibited. The necessity to remember characteristics of operators, such as limits on their applicability, or consequences of applying operators is eliminated. A way to implement some degree of

Externalization is to make parts of the interface context-sensitive, e.g. by hiding or disabling functions that are not applicable at the moment. By doing so, the user is “taken by the hand” by limiting choices and providing feedback (Van Oostendorp & De Mul, 1999). Examples are wizards, help-options and grayed-out menu-items that do not permit using them, thus offering a context-sensitive interface where only possible actions are displayed. For example, in MS-Word, one cannot select “paste” from the menu when nothing is copied first. “Paste” shown in gray color indicates *both* the presence *and* the unavailability of the command. “Graying out” *externalizes* the applicability of an operator at a certain moment, and makes memorizing this information unnecessary (Figure 1.1).

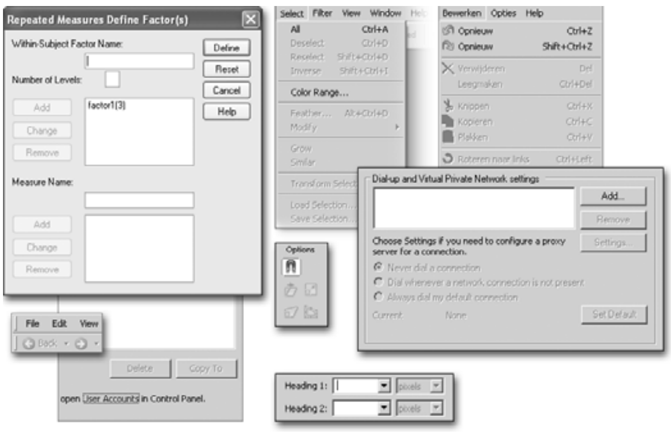



Figure 1.1 - Externalized information in MS-Windows applications

In the opposite situation, when no such features are provided, a user has to *internalize* the information by himself, and store this information in his memory. The above definition of usability implies that systems are tools we use to accomplish a certain goal (text processing, spreadsheets, editing video, looking for information on the web). Rubin (1994) was probably right, though it can be interesting to look at what these presumed friendly systems actually do to humans and their behavior on a deeper level, and whether the highest achievable usability possible in the classical sense of the word is in fact *always* desirable.

## 1.2 Beyond usability: Provoking deeper contemplation and attention to the task

Usability engineering is extremely important, the guidelines and techniques are improving many interactions with systems, and we can now reap the harvest of years of research. The usability industry is indispensable and crucial to the success of a broad range of applications. Many usability guidelines and practices help designers negotiate between the complexity and limitations of technology and human nature, in trying to make interactions as close as possible to how we are. Think of the daily life metaphors that are being used in GUI’s and how developers attempt to adapt systems to the mental models we have of the everyday world. What is better than buying an online book or a train ticket within one minute, or having a PC that our grandparents can also use? Thanks to navigational GPS devices in vehicles such as TomTom ™, we will never have to unfold a map again, but simply follow “go left” and “go right” instructions instead of risky escapades such as driving and looking at a map at the same time. Although not so much of interest here, one could wonder whether our once developed navigational skills might deteriorate after years of use. Still, it is certain that some activities are easier and sometimes more pleasant than before.

If one takes a broader perspective one could go back one level to where we started, and realize that usability was born from the discipline of "Human Factors" (also known as ergonomics), which has its roots in psychology. The methods and practices originated within the US military during World War 2. Sophisticated weaponry had to be usable, because if it could not be used, military objectives would not be met (if it was used incorrectly, using it would kill the forces rather than the enemy). Human Factors research concerns the study of human beings and attempts to understand the advantages and limitations of the human body and mind, and how human performance is affected by environmental factors. When applied to more specific contexts, human factors research could for example focus on *general* human behavior in relation to technology (such as studying how people perceive fonts, colors and links in a webpage). Most of the usability guidelines are quite general. However, human factors research can also focus on more specific environments, situations, tasks or products, depending on the objectives. The fact that computer systems are friendlier and more ubiquitous every day (also thanks to usability research) does not necessarily mean that implementing usability guidelines causes us to do *all* computer-based tasks "better". The reason is that "better" strongly depends on the goals that the users of the system have. Regarding the almost ubiquitous use of computer systems today, the goals might in some situations have to be extended beyond seemingly friendly machines and systems that "understand" humans. One can think of situations where learning itself is the aim, or where it is important that at all times users have an active command of underlying rules.

Outside the world of research, one can sometimes encounter unscientific notions (some more grounded than others) reflecting that many computer systems of today "make us stupid". An often-heard remark is that in the days of command-line interfaces such as MS-DOS or UNIX, people had a much better knowledge of what it was they were actually doing, and had more control over the system. To execute actions, users had to know exactly which command to type, and this knowledge had to come from the memory of the user. Commands had to be learned first and had to be stored in a person's memory to be retrieved at a later moment. Another thought is that the direct manipulation environments that make error recovery so easy these days, and the fact that anything can be dragged and dropped with the mouse, cause users to have a different attitude towards their tasks. Making mistakes often causes little or no harm, and a "let's see what happens" attitude can prevail. In addition, "mousing" is perceived as less work (less cost) than typing commands, both cognitively and physically.

Concerning the issues of the cost of actions and the level of planning that takes place during task performance, O'Hara and Payne (1999) conducted several experiments (see also section 2.5.2). They examined the hypothesis that the cost of performing an operation affects planfulness and concluded that planfulness in problem solving behavior can reliably be influenced by the implementation cost of an operator. When the cost associated with an operator was relatively high, problem solving strategies became more *plan-based*, whereby search paths were considered and evaluated mentally. Conversely, when the cost associated with operators was relatively low, subjects shifted to a less planful strategy in which search was more reactive and *display-based*. The different levels of planfulness also had differential effects on learning performance. Subjects trained using an interface with high operator implementation cost, performed significantly better on subsequent problems in the same domain (in terms of shorter solution lengths and decreased completion times) than subjects trained using a low operator cost interface.

In the context of users who have exact knowledge of what they are doing, it is interesting to look at the use of menus. A way to accommodate hundreds of commands in a GUI is by nesting commands in menus, submenus, and sub-submenus. This is necessary because of screen space limitations, if one wants the interface to be manageable by mouse interaction and avoid typing command lines. Essential in this situation is that commands do not (completely) have to come from the memory of a user; mostly a vague recollection of where a command can be found suffices. In other words, commands can be chosen based on "recognition", rather than on the basis of "recall", as was the case in command-line interfaces. This is often very convenient, and allows a very broad range of users to use applications in an easy manner, so the usability target has been achieved in that respect. However, Mayes, Draper, McGregor and Oatley (1988) investigated what users remember of the detailed content of the Mac Write interface. It showed that even experienced users recall little of menu contents, even though they use them all the time and achieve successful performance. Although the needed information for the task was found and applied, it was not remembered later on.

O'Hara and Payne's manipulations to make interactions more costly in their experiments, can be seen as making the system more unfriendly. Another way to purposely make a system not as friendly as it could possibly be, is to do something as unintuitive (regarding today's processing power) as making a system slower than necessary. This is exactly what O'Hara and Payne (1999, see also section 2.5.2) did in another experiment. In their tasks, they investigated the effect of lockout time. Lockout time meant that in the task, after a move, subjects had to wait for a beep that indicated the next move could be made. This made interactions with the device in the case of errors a time consuming, costly endeavor. The idea was that perhaps subjects want to avoid longer waiting times and are encouraged to pay more attention to and think harder about the results of interactions. This was indeed the case, in line with the predictions of the cost-benefit model of planning/action levels. The experiment showed that a longer lockout time resulted in shorter solution lengths.

The general idea that can be extracted from the above examples is that a friendlier interface does not necessarily result in better performance, but depends on what one aims at. Of course, in many cases, *not* exactly knowing which commands are where does not cause any problems. In addition, the interesting finding that a more unfriendly system can provoke better performance does not mean we have to make our systems slower. The examples above are meant to indicate that there might be slightly more to consider when making choices in systems design, and instilled the motivation to look deeper into the nature of interaction in certain situations and tasks. Users not being permanently in command of what function can be found where (or even know of its existence) or the notion that a friendlier system can result in more shallow behavior, are findings that are worth delving into. This becomes interesting especially when the goals of users doing a computer-based task are of such a nature that not knowing or finding the right command as fast and directly (efficiently) as possible can have severe consequences. These issues are important in safety-critical systems or in situations where knowledge construction *itself* or transfer of skill is important. It can be worthwhile to take into account that not all interactive systems have *only* the goal of “ease of use from the first time on” and “maximum user satisfaction”. These notions by no means deny the fact that any system needs to have some degree of usability in the first place; if not, one could not work with them at all. However, one can think of various situations where the goal is beyond plain usability. There are situations where 100% command of all functions and continuous proactive thought and planfulness are essential. Examples where system designers should *not* neglect this because it can have unwanted or potentially disastrous consequences exist in various contexts, such as:

1. Contexts where software is used to convey knowledge or provoke exploratory behavior from users (Moreno & Mayer, 2000). In education, software is used to visualize concepts, explain theories in an interactive way, whereas earlier we had to read them in books and then put it together in our minds. It would be probably more pleasant and appealing to have an interactive animation of how the human heart functions, than having a static image in a book and having to understand and mentally visualize the process. Nevertheless, in the latter case, because deeper thought processes from the learner are required, learners perhaps store the knowledge more solidly, exactly because they did more processing themselves (see also Van Oostendorp & Beijersbergen, 2007).
2. Applications in the medical domain, where decisions and actions are of such a crucial nature, that making errors has severe consequences (Koppel, Metlay, Cohen, Abaluck, Localio, Kimmel & Strom, 2005). For instance, a medical recommender or expert system can be very helpful to obtain a second opinion, but is there a risk that users rely too much on these systems? These systems are meant to help, but could there be the danger that users assume that the work is being carried out for them? It is extremely important that users are at all times proactive in their thinking, and maintain 100% command of what underlies a situation.
3. Safety critical systems such as air, subway or train traffic control, industrial processes, vehicle navigation, hoisting cranes and so on (Boucheix, 2004).

We suggest that the usability field is perhaps too much subject to local theories, rather than a theory of interaction based on cognitive principles. Therefore, this research focuses on deeper aspects of human computer interaction. Our general idea is that in certain situations, if a system is less (seemingly) friendly, it could be beneficial for task performance (O’Hara & Payne, 1999). By forcing users to actively internalize concepts and principles that underlie a task or domain could result in more efficient performance, and more solid knowledge. Our aim is to propose an interaction framework with a trade-off between common usability on one side (designing systems so they are pleasant to use, externalize information to aid users) and designing systems such that they continuously keep the user proactive (forcing users to internalize, avoid lazy user behavior at all times) on the other side. The issues we focus on will be described in the next section.

## 1.3 Our research issues

Our research in essence (on a user-machine level) concerns the amount and the way information is visualized, the way systems provide feedback, and how users are guided through problem solving processes. Implementing usability is often focusing on letting systems behave the way we humans are, and adapting them to our mental models. However, it is also interesting to turn this around, which would result in “how can we let humans behave the way we want with our systems?”. Of course, one should not take this the wrong way in thinking that a “system” wants us to do things, or that humans should adapt to machines at all times. What is meant is what is behind; what was the goal of the software: it should “help us do” something, but we also sometimes still want users to behave in a certain way, while not losing effectiveness. What kind of (cognitive) behavior do we want to provoke in users, depending on the task? In critical tasks, it can be desired that a user is constantly and consciously in control of planning and the selection of strategies, which in turn influences performance while doing a task. This is referred to as *metacognition*, a level of thinking that involves active control over the process of thinking, planning the way to approach a task, monitoring comprehension, and evaluating the progress (Ridley, Schutz, Glanz & Weinstein, 1992). If one wants to provoke metacognitive behavior, can the interface with which we do our task have an influence? With regard to externalizing information to relieve cognitive load, a potential danger of this technique might be that users do not need to reason for themselves. They can be seduced to neglect to look for underlying rules, processes and other information while this is in fact necessary in order to build stable knowledge structures (schemas and plans) that can also be applied in new situations. In addition, if Externalization of information (which might feel friendly) *does* have the effect that users reason less themselves, does this have its repercussion on performance, memory of the task and underlying rules, or the solidness with which it is imprinted in memory? Moreover, if the kind of interface affects the long-term imprinting in memory, or quality of skill we acquire, does this have an effect in situations where transfer of skill is necessary? In accomplishing the goal the software was intended for, there might be a trade-off between on the one end plain usability as measured by perceived ease of use and user satisfaction, and on the other end designing a system in such a way that active contemplation and mastery of what is underlying is provoked. By this trade-off, we mean that an adequate balance has to be sought after. While attempting to provoke mental alertness, at the same time a system should not be too hard on the user (the task should not become a burden). On the other hand, however, a system should not be too inviting to act shallow and lazy because the user can have the feeling the task is being carried out for him (at least to some extent).

Perhaps the display-based characteristics of a device (e.g. Payne, 1991) which permit rapid and low-effort input do not encourage reflection as much as they should. It might be that the current generation of computer users (even if they are only occasional users) has learned that it is highly unlikely that their actions will bring about irrecoverable damage to the process or system. In modern GUI’s, users often click around without control, clicking and clicking on and on has become our second nature, perhaps partly due to the World Wide Web. The *undo* facilities that became almost mandatory in systems are mostly good and handy features, but perhaps they changed our ideas about the consequences of our actions in computer environments. A section in a book chapter by Bannon (1991, p.28) focuses on allowing for active users, stating:

*“While focusing attention on the user may be a positive step, users are not simply passive objects that others must study and design for, as some accounts would have it. People are,*

*or can become, active agents. They often wish to accomplish tasks, to understand what is going on, and are willing to jump ahead and explore the computer…".*

Still, while assuming people possess this activeness, one has to delve into human nature, and be aware of the mechanisms that are natural to us, which include, unfortunately, laziness and shallow behavior. There is nothing we like better than the feeling that things are being done for us, and in many cases, this can be exactly what is desired. However, in some situations one has to be careful to neglect these characteristics. This research concerns exploring the conditions under which motivation and deep contemplation for the task can be provoked to achieve better task performance, fewer errors, and having users that constantly are on top of the task.

In the literature, a differentiation is made between *plan-based* and *display-based* problem solving. The first type is characterized by having to learn and use detailed plans for solving the problem, necessitating *Internalization* of all needed knowledge. This knowledge gradually builds up in Long Term Memory (LTM) through reflection and in making inferences, and is self-derived. These processes make large demands on Working Memory (WM) and LTM, at least during learning. We will elaborate on these processes in chapter 2. The notion of plan-based behavior is reminiscent of what Rasmussen (1983) referred to as knowledge-based attentive action, which uses internally formed mental models, demanding high workload.

Display-based problem solving, on the other hand, uses information available on the interface to structure and guide problem solving. We call such information *externalized*. Little WM and LTM involvement ensues, and only recognition, not recall, is needed for task performance. Both recognition and recall involve retrieval from LTM, but recognition is a much less intensive memory process than recall. Several research outcomes and guidelines (e.g. Zhang & Norman, 1994; Zhang, 1997; Apple, 1992) point at the advantages of display-based interaction (see also section 2.5.1), but based on other research (O'Hara & Payne, 1998, 1999; Trudel & Payne, 1995), we hypothesize that there are also costs involved. Task relevant information being visualized by a system, and system feedback, is henceforth referred to as the *Externalization* of interface information. This is in contrast with *Internalization*, where this is not done (or to a lesser degree).

This project investigates the conditions under which *externalizing* interface information by interface controls, influences a user's performance in solving problems requiring planning. Planning is a metacognitive activity which can be expected to be crucial for the formation of a mental model of the problem to be solved, and consequently for task performance. We chose this category of tasks, because there are indications that a strong reliance on external information leads to a negative effect with regard to planning of behavior (O'Hara & Payne, 1998, 1999), while this might be the key to success in problem solving. We aim to specify which interface presentation (with or without Externalization) leads to the best performance when learning to do a task, when performing the task after learning, and when performing it under altered circumstances, and to the best long-term memory of the task. The effects of leaving out or varying the amount of task information in an interface (which has become default by now in a wide range of applications) will be studied. Furthermore, we will pave the way for the notion of *information gaps* (see section 6.4.1). This term refers to the fact that certain information that could be seen as default is purposely left out. This can be seen as corresponding with Internalization, purposely *not* providing information that could be easily made available otherwise. The idea is that by doing this, a user will be *even more* motivated to think hard, and apply deeper cognitive strategies than when this information would be there. In summary, our research issues can be broadly placed on three levels:

- **Behavioral level**
  How does interface style (Internalization or Externalization) influence user behavior and performance in tasks requiring planning, on performance when learning to do a task, performing the task afterwards or under altered circumstances, and long-term memory of the task?
- **Theoretical level**
  By answering the above question, we intend to obtain more insight in how, and to which extent plan-based vs. display-based problem strategies are provoked by the interface style. We try to gain insight in mental processes involved, and how we can connect these to existing theoretical notions.
- **Practical level**
  Insights obtained by answers on the above questions can help us to contribute to a predictive theory (interaction framework) about when and where to apply Externalization features (or not) in the interface in order to provoke more display-based, or more plan-based approach.

Results of the research can lead to guidelines for design of human-computer interaction for optimal problem solving. It answers questions such as: does visual support (Externalization) in problem solving help, under which circumstances and in which way. It aims at contributing to the theory on how knowledge organization, learning, and memory relate to modern insights in computing and visualization in human-computer interaction. Our aim is to propose an interaction framework with a trade-off between common usability on one side (externalize information to aid users), and designing systems in a way that keeps the user proactive on the other side (avoid shallow processing). As such, it contributes to more insight into cognitive processes and aims to result in improvement of the interaction between man and computer.

## 1.4 Primary tasks and secondary tasks

To illustrate our issues from the preceding sections and to obtain a better understanding of what actually can happen during human computer interaction it is useful to emphasize the difference between primary tasks and secondary tasks when using a computer. One can look at producing some piece of text in for example Microsoft Word ©. The task itself has several non-computer equivalents such as writing with pen and paper, or more sophisticated equipment, such as using a typewriter. In the case of a typewriter, one has to have knowledge of how the machine works (how to produce uppercase fonts, how to insert space between paragraphs, and how to change the color of what is typed by changing the ribbon. On the computer, one can do the same thing, already with many more functions, especially regarding formatting, but what is needed is a way to access these functions. This is communicated by the interface, and has to be learned, at least to some extent. A lot can be done thereafter on the basis of recognition: to look up the desired command, often a vague recollection of where a command can be found suffices. The *primary task* (producing text) is the same with handwriting, a typewriter, or a text processor, only the tool is different.

Managing and using the tool is the *secondary task* that is needed to perform the primary task. Of course, there are still other factors at play (experience, time pressure etc.) that have their influence on task execution and consequently the final output (see Figure 1.2).

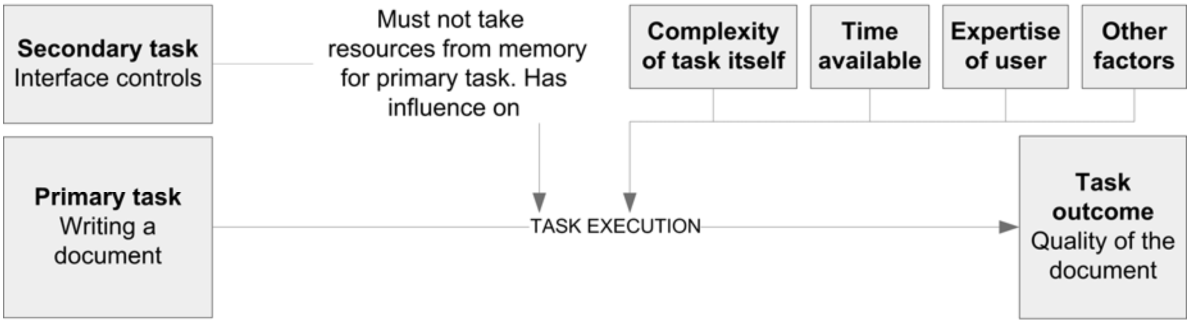



Figure 1.2 - Primary and secondary task

In computer tasks, we prefer to concentrate on the primary task itself rather than on the interaction with the interface, so the higher the learnability, smoothness, intuitiveness and thus the usability of the secondary part, the more efficient and satisfying performing the primary task will be. The secondary task, managing the interface, will (hopefully) have no influence on the quality of the text written, it is only a means to accomplish the primary task (Van der Veer, 1990). Of course, this is not always the case, and the statement that computers often seem to complicate life instead of simplifying it is not unheard of. If the tool for the secondary task is badly designed, it *will* have an influence. When the secondary task is cumbersome and time-consuming, it will lead to less time and effort devoted to the primary task and a frustrated user. This will have an effect on the quality of the output (the primary task). This is often seen in business, and reported on in an interesting book by Landauer (1996), who investigated the link between information technology and productivity and showed that it did not live up to its promises. However, for the sake of the explanation, let us assume that our secondary task (the interface) is easy enough to use and interference with the primary task is reduced to a minimum. For example, when the earlier mentioned issue of graying out menu items takes place and one cannot click “paste” because nothing was copied first, this informs the user of the secondary part, and one could assume that this is a good thing. What is communicated to the user is the flow in which things are supposed to happen, and this information does not necessarily affect the quality of the primary task: the text that is produced.

To illustrate the distinction between primary and secondary task and to obtain a clearer idea of what actually happens during interaction regarding primary and secondary task as we see it, one can take the example of chess. Chess has an old-fashioned way to play it, but also many modern digital equivalents. The rules of the game are not so difficult to manage (but being a great chess player is something else). One can play it on a board, but also with a different tool: the computer. Let us have a look at the rules that apply to a horse in chess. When nothing is in the way, a horse can move to a maximum of eight different positions from the square it is currently at. It can jump as follows:

- 2 squares forward **or** backwards > 1 square left **or** right
- 1 square forward **or** backwards > 2 squares left **or** right

Imagine a user who does know how to play chess has to take place behind a chess computer and is asked to make a smart move with the white horse on the left in situation in Figure 1.3.

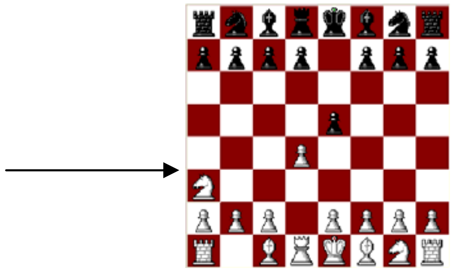


Figure 1.3 – Chess

The interaction between system and user when the primary and secondary tasks are completely separated is depicted in Figure 1.4. The secondary task (the interface) plays a minor role, it is only the way (the tool) to move pieces with, and is not expected to interfere much in the thinking process and affect the quality of the move.

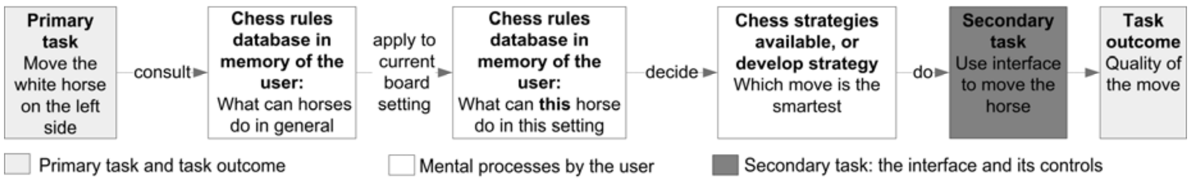



Figure 1.4 - Separated Primary and secondary task

However, there are different situations where the distinction between primary and secondary task is much more blurry. Let us again take the example of chess software. To learn the game, or to practice, a lot of chess software applications, such as Fritz © (2003), offer *training modes* in which exactly this blurring between primary and secondary tasks occurs. In the figure below, when it is white's turn and one clicks on the white horse on the left side, the rules for the horse to move are *externalized*. The legal target squares are highlighted in green, thus informing the player of where the horse can go; three squares are possible here (Figure 1.5).

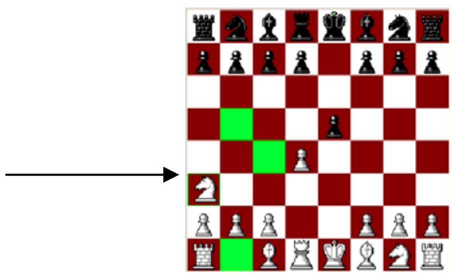


Figure 1.5 - Chess with guidance: externalized information

Here, the primary task and secondary task become intertwined. This happens because the interface controls (the secondary task) become part of the primary task (moving the horse) in the sense that it takes over steps that otherwise are (mentally) carried out by the user. When the primary and secondary task are partially integrated the interaction between system and user changes. There is an essential difference between the flow of events in Figure 1.4 and 1.6. In Figure 1.4, the second and third boxes, which are mental steps, are being done by the system and are externalized. These two steps become part of the secondary task, and cause the user to see the legal actions that can be performed in a split second. In this situation, mental consultation concerning the rules of the game in general and the specific rules that apply in this constellation of the board has to be done.

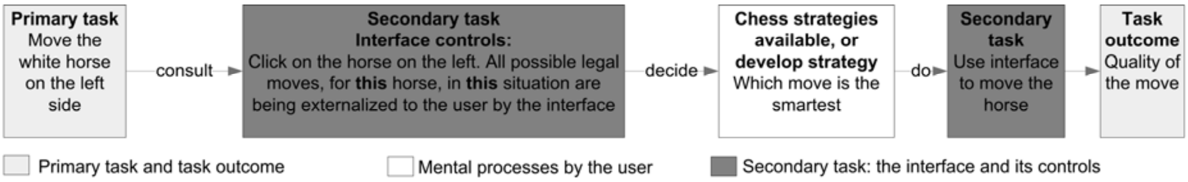



Figure 1.6 - Partially integrated primary and secondary task

What actually happens in the example in Figure 1.5 (when the primary and secondary task are partially integrated) is quite common in modern GUI's, and is referred to as *destination feedback*. Sun Microsystems (2001), in their guidelines regarding drag-and-drop situations, state: *"Provide destination feedback so users know where the dragged object can be dropped. Highlight drop targets when the pointer is over them to indicate that they can accept the selection or source"*. It is this kind of Externalization of information that we study in this research project. One can think of several positive and negative effects regarding this Externalization of information in the chess task:

**Possible positive effects of this Externalization of information**

1. **Computational offloading.** Because several mental processes are skipped and a mental visualization of moves is done for the player, the cognitive load on the user is lower (Scaife & Rogers, 1996). The user can devote a maximum of mental resources to other processes (such as developing a winning strategy).
2. **Suggest moves the player did not think of**. Seeing all the legal moves at a given moment with a given piece could lead the player to a move that he/she by himself would not have thought of.
3. **Learn chess rules.** Because of this continuous information, one could expect the player to learn by heart what chess pieces are allowed to do in general.

**Possible negative effects of this Externalization of information**

1. **It can provoke laziness and shallow behavior**. This assistance could provoke laziness, shallow behavior and an overall lack of focus because of the feeling that "the work is being partly carried out for the user". It shows legal moves, but this information by no means tells you what the best or smartest move is. Perhaps the Externalization hinders metacognition or active control over the process of thinking, and this is exactly what is needed to improve the problem solving process.
2. **It does not trigger contemplation of the rules and does not aid strategy construction**. The user is not occupied with inference and storage (in the case of a new task) or retrieval and application of rules because step 2 and 3 of Figure 1.4 are omitted.

Furthermore, playing a good game of chess involves not only local analyses like this one with the horse, but large series of sequences. Developing and applying a winning strategy does not depend on one move. It is known that good chess players think ahead many moves (including those of the opponent). In the thinking ahead process, mental visualizations of moves in the future are made without assistance. This assistance of Externalization will probably not serve a better overall strategy. The externalized information is not necessarily obstructing adequate strategies, but it is certainly not fostering active contemplation.

3. **Distraction**. While assuming that some mental relief takes place, in the sense that step 2 and 3 of Figure 1.4 are skipped, the interface information in the form of green squares lighting up might have distracting properties. The interface information still has to be processed. If the player is studying the board and watching the possible squares, several of them can make absolutely no sense. However, the user cannot ignore to see them, and attention will be drawn to the green squares.

The purpose of this particular example is not to discard the usefulness of this kind of functionality in software in general. It is meant to illustrate what can actually happen when software externalizes task specific information, and to pave the way for our research on Externalization of information in computer-based problem solving. Leaving *out* information that is meant to help, thus making the system less user friendly (at first glance) could have beneficial effects in the end.

## 1.5 Research questions

Following our research issues (section 1.3), our research questions fall into 3 levels. The first level concerns how interface style (Internalization or Externalization) influences behavior and performance in tasks requiring planning. The second level concerns which theoretical concepts and frameworks we can position our findings in. The third level concerns how our findings can contribute to practice and involves constructing an interaction framework about when and where to apply Externalization features (or not) in the interface.

### 1.5.1 Behavioral level

We will investigate the conditions under which *externalizing* interface information influences a user's behavior and performance in solving problems requiring planning, be it positively or negatively. Our main research questions are:

- In tasks where planning and deeper contemplation is required, which interface style leads to more plan-based behavior, better strategy, and consequently better task performance?
- Besides immediate performance, which interface style causes the best knowledge about the tasks, rules and solutions afterwards (shortly afterwards, and after a longer delay)?
- Which interface style leads to the best performance in a transfer situation (altered task or interface circumstances) and better resistance against a severe task interruption?

Our assumption is that the Externalization of information even though fewer rules have to be learned and retrieved from memory, leads to shallower processing, less proactive contemplation and less metacognitive activity. Because of this, we expect Externalization to result in worse performance (less efficient solutions) on problem solving tasks. An interface where users have to internalize certain information by themselves without guidance, might seem like a less user-friendly system, but is expected to result in more efficient solutions and better imprinting of knowledge in memory. To answer our research questions, and to see whether our assumptions are valid, we investigate the following specific questions which we wish to provide empirical evidence for:

1. **Repeated problem solving tasks. How does problem solving performance evolve over time?** Different groups of users will perform repeated problem solving tasks with one of the two interface styles: Internalization and Externalization. We expect that in the very beginning, Externalization can have some advantages, but that these will disappear quickly and that the Internalization subjects will outperform the Externalization subjects.
2. **Transfer of skill to another task. Does learning and doing a task in one interface style have an influence on performance for a similar (but different) task?** We expect that when one has learned and performed a task with an interface style that lacks *externalized* information (requiring a lot of Internalization from users), metacognitive activities such as planning are enhanced. This will result in better strategies and better imprinting of task knowledge. We expect users who worked with the Internalization interface are better able to transfer and apply this task knowledge to different, related tasks.
3. **Transfer of skill to another interface style. Does learning and doing tasks in one interface style have an influence on performance of that same task afterwards with another interface style?** We expect that if one has learned and performed a task with an interface style that *has externalized* information, doing that same task with the interface that *lacks Externalization* (Internalization) will cause deteriorated performance. This is because users first partially relied on the guidance by the Externalization interface, and their internalized knowledge will be suboptimal. When this guidance is suddenly missing, they will perform worse. When it is the other way around, from the *Internalization* to the *Externalization* interface, we expect the same performance since users are already used to do the task based on information processing in their own memory.
4. **The influence of instruction to the user before the tasks. Does instructing users to plan vs. not doing this influence behavior, and is this different in the two interface styles?** We assume that the Externalization interface discourages planning and contemplation. It might be the case that instructing users to "*plan and work concise and efficiently*" causes users in the Externalization version to put in more effort, and that the performance differences between the Internalization and Externalization interfaces decrease. However, when subjects are instructed to do it "*as fast as possible, making mistakes is not a problem*", Externalization subjects might perform even worse than without instruction. This might affect Internalization subjects less, because they are expected to be continuously triggered to plan and think by the interface.
5. **Individual differences in cognitive style. Do individuals who like to engage in effortful cognitive tasks behave differently, and is this different in the two interface styles?** Above we focused on externally induced motivation towards the tasks. However, we will also focus on more stable personal qualities of people that could influence behavior. We will measure the tendency of individuals to engage in and enjoy

effortful cognitive tasks. All people need to make sense of their world (e.g. in a computing environment), but they do this differently. Some persons love to look for, reflect on and reason about information, whereas others only think as hard as they have to and are inclined to rely on others. We expect persons that love to look for, reflect on and reason about information to be less affected by the different interface styles.

6. **Delay: Do the two interface styles have different effects on performance after a large delay?** Since we expect that the externalized interface style provokes less solid imprinting of rules and strategies, we expect these users to have worse memory for the task, its rules and solutions than the Internalization subjects. We will re-test the same subjects several months after first doing the tasks for the first time.
7. **Interruptions in the problem solving process**. **Do interruptions have the same effect in the two interface styles?** If users have learned the task by doing it several times in the condition where they had to internalize information themselves, we expect better command of and better imprinted knowledge about the task, rules and solution procedures. We will impose an interruption of several minutes that is severely disrupting on the users in both the Internalization and the Externalization condition. Our assumption is that if indeed knowledge and procedures are better in the Internalization condition, this will result in smoother and more efficient task resumption for the users that had to internalize all the information.

### 1.5.2 Theoretical level

The specific questions mentioned above all involve finding answers to our main research questions. By interpreting the answers to these questions, we intend to obtain more insight in how, and to which extent plan-based vs. display-based problem strategies are provoked by varying Externalization vs. Internalization in the interface style. We try to gain insight in the mental processes involved and how we can connect these to existing theoretical notions that are important in our research (see next chapter), such as Cognitive Load Theory and Information Processing Psychology. Besides, we aim to contribute to the larger body of theory in general concerning representations in HCI, and specifically in interfaces.

### 1.5.3 Practical level

If it is indeed the case that users that have to *internalize* certain information show more plan-based behavior (reflected by better strategies, higher efficiency, better knowledge about the task and rules, better performance after an interruption and in transfer situations), we can accommodate the findings in an interaction framework. This interaction framework in turn, can be valuable for the HCI field, especially when it concerns systems where learning itself is the aim, or when the tasks are of a critical nature where interactions or making errors comes with a great cost. Each of the above specific questions in section 1.5.1 will be accommodated in the interaction framework, which can be used to guide design decisions concerning interactivity and how controls and operators in the interface should convey externalized information.

## 1.6 Research Method

Throughout the research project, controlled experiments are conducted using two interface styles: one version in which certain information is externalized onto the interface (Externalization) and another version where this is not done (Internalization). What in essence will be varied is the degree of interactivity, which we believe can have an influence on cognitive behavior (associated with motivation to plan, and the willingness to make inferences). In the Externalization versions the operators in the interface convey information (and thus become part of the problem space), in the Internalization versions this is not the case. All the experiments, except for experiment 5, are conducted in the Usability Lab at the Center for Content and Knowledge Engineering at Utrecht University. The first series of experiments makes use of a fairly abstract task: a computerized isomorphic version (in several variations) of the well-known Missionaries & Cannibals problem. The second series of experiments uses a more realistic office-like application we developed called "Conference Planner", a constraint-satisfaction scheduling task. Several measures will be collected, among which: path efficiency, routes chosen through the problem space, type of strategy applied, and timestamps. We will also assess knowledge and study various mouse-related events and eye tracking data. These measures will be obtained by computer-recorded log data, eye tracking data, and questionnaires.

## 1.7 Overview of the dissertation

In this chapter, we introduced the scope of this research and our research questions. Throughout the research, the central issue is which effects Externalization information in the interface has on various measures of performance. Our questions can be answered on three levels: a behavioral level (performance), a theoretical level, and a practical level. Our research questions will be empirically investigated, by using two interface versions: one where certain information is externalized and another one in which this is not done. The remainder of this thesis is structured as follows:

Chapter 2 presents the theoretical background to Chapter 1. We will elaborate on several theories and notions regarding cognitive processing in problem solving situations, among which Information Processing Psychology, Cognitive Load Theory and theories of Distributed Cognition and External Cognition. We will mention relevant research concerning computational offloading and cost-benefit analysis, in order to arrive at our own approach in this research. After that, we will arrive at the theoretical framework we adopt, which will primarily be connected to the notions of Cognitive Load Theory. We end the chapter by mentioning the choices we made in our experiments, and by explaining how we will implement the technique we propose: Externalization of information by destination feedback.

In Chapter 3, we present the empirical results of experiments 1, session A and B, and experiment 2. We begin by elaborating on the theoretical background of the applications we developed: computerized versions of a famous math puzzle (Missionaries & Cannibals), which we called Balls & Boxes. In these first experiments, always two different interfaces to the same task will be used, and we analyze differences in problem solving performance. The focus in experiment 1, session A, is on general performance on the tasks, and performance after a pause in which a distraction task takes place. Experiment 1, session B, is a rerun with the same participants as session A, but months later, and the aim here is to see how task information is remembered, and how it is applicable to a transfer situation. Experiment 2 also concerns a Balls & Boxes experiment, but has minor variations, and focuses on the influence of explicit instruction to plan to seduce subjects into certain behavior.

In Chapter 4, we elaborate on experiment 3 and 4, which use material that is much more realistic and life-like. These are constraint-satisfaction scheduling office-like tasks, with applications called "Conference Planner" and "Ferry Planner". We begin by explaining our choices in material construction and method, followed by the experimental results. The focus in experiment 3, besides general performance measures, is the influence of personal attitudes and characteristics that are already present in people before the experiment, and how this relates to behavior instigated by our different interface versions. In experiment 4, we focus on the influence of a severe interruption, on transfer of skill to another task, and to an interface version the subjects have not used before.

Chapter 5 elaborates on experiment 5, which uses the same applications as experiment 3 and 4. Here we have three research foci. The first is obtaining more insight in transfer than in the preceding experiments, and the second is obtaining better insight in the strategies subjects use, and how this differs, instigated by the different interface versions. The third focal point concerns obtaining additional backup for our assumptions, by applying eye tracking techniques to see whether eye fixation patterns coincide with the behavior and performance we assume (depending on the interface version).

In Chapter 6, we discuss the findings of the five experiments. We connect our findings back to our main, and our specific research questions from Chapter 1. We discuss our findings on three levels: a behavioral and performance level, a theoretical, and a practical level. On the behavioral level, we resume and discuss the results of the experiments. On the theoretical level, we discuss our findings in the context of our theoretical framework. We discuss whether our theoretical choices and assumptions were feasible, and connect back to research findings of others. On the practical level, we discuss the implications of our findings for the HCI practice. We present some thoughts on how our contributions may be valuable to the HCI field, and propose an interaction framework. The chapter ends with thoughts on how research fields like ours and the field of game research can mutually profit from each other.

In Chapter 7, we firstly draw conclusions about our findings, again on a behavioral and performance level, followed by conclusions on a theoretical, and a practical level. Lastly, some directions for future research are presented.

# 2 Theoretical background

In this chapter, we will elaborate on several theoretical frameworks that are connected to our research issues, and mention some more findings from relevant literature. In the previous chapter, we have laid out our research questions and general method, and elaborated on the concepts of the Internalization and Externalization of information and on how to instigate planning during problem solving. There have been various researchers who analyzed the interaction between internal and external representations. A central issue in our research, is how plan-based behavior can be provoked (some of them were already briefly mentioned). The idea of planning in which complete action sequences are computed in memory before enacting them in the world, has been criticized in light of evidence from studies of human problem solving. We will briefly mention some relevant frameworks from which HCI and the presentation of information are being studied. They are not mutually exclusive, there is overlap between them, and we will mention several in order to refer to them later on when elaborating on our own findings.

## 2.1 Information Processing Psychology

Computer-based problem solving situations such as the ones that will be focused on in this research are often studied using notions from Information Processing Psychology (Newell & Simon, 1972). Information Processing Psychology (IPP) seeks to explain cognitive processes during various kinds of human thinking, learning, and problem solving, and the central claim is that humans operate as an information processing system. It is a general theory of human cognition, which assumes that thinking is information processing, in which the mind is seen as a system that processes information through the application of logical rules and strategies. In IPP, cognition is largely regarded as a type of computation. When an individual perceives, encodes, represents, and stores information from the environment in his mind, or retrieves that information, he is thinking. In IPP, task environments are described in an objective, abstract manner, which could in principle be applied to any problem solving situation. Newell and Simon (1972, p.810) proposed that problem solving consists of a search in a problem space (see also section 3.1). A problem space consists of:

1. A set of elements, U: symbol structures, each representing a state of knowledge about the task
2. A set of operators, Q: information processes, each producing new states of knowledge from existing states of knowledge
3. An initial state of knowledge Uo: the knowledge about the task the problem solver has at the start of problem solving.
4. A problem, which is posed by specifying a set of final, desired states G, to be reached by applying operators from Q
5. The total knowledge available to a problem solver when he is at a given knowledge state

According to IPP, a problem consists of a gap between an initial state (U) and a goal state (G). The means of solving this problem involves applying an appropriate set of operators required to complete a series of state transformations that will eliminate the gap. These transformations must be accomplished without violating any of the conditions on the operators. IPP suggests several stages in problem solving which would generally look as follows:

1. Identify the problem space and the initial and goal state. These two states define the boundary of the problem space. The larger the "distance" between the two states the larger the problem space.
2. Identify the intermediate states between the initial and goal state.
3. Identify what needs to be done: the "moves", which enable the problem solver to get from one state to another. In order for a problem to be solved, there has to be some procedure by which the situation is transformed from one state to another.
4. Identify the resources, e.g. knowledge and skills needed to execute each of the moves.

Information processing theorists see memory systems as functionally separate systems and see human memory as divided into different parts. During processing, information from our senses is initially stored in *Sensory Memory* (SM), a memory buffer holding sensory input. The next component that differs structurally and functionally is *Short Term Memory* (STM). Nowadays, and henceforth in this text STM will be referred to as the already mentioned term Working Memory (WM). Another component is *Long Term Memory* (LTM) where (in theory) an unlimited amount of information can be stored. The longer information is held in WM, the greater the chance that it will make it into LTM. Information from WM can pass to LTM through the process of rehearsal and meaningful association. In LTM, several kinds of knowledge are stored, such as:

- Declarative knowledge: memories that are consciously available. It can be subdivided in
    - Episodic memory. Memory for specific events in time, e.g. year of the first moon landing
    - Semantic memory. Refers to knowledge about the external world, the memory of meanings, understandings, and other concept-based knowledge unrelated to specific experiences, such as the fact that a dog is a mammal.
- Procedural knowledge: how to ride bicycle, how do I make coffee

In Human Computer Interaction research, investigators often resort to IPP, since interaction with some system often is fundamentally an information-processing task (Card, Moran & Newell, 1983). Human behavior, from displayed information to an eventual action response, takes place in several processing stages, and theories about information-processing capabilities (attention, memory, decision-making) are taken into consideration in designing interfaces and tasks. Concerning human capabilities and the nature of human memory in connection with interface design, it is clear that the brain has a limited capacity for the amount and nature of the information it can process. When designing systems one needs to take into account that the bottleneck that can occur in sensory and working memory has an influence on how we process and learn.

An influential approach stemming from IPP is *rational analysis*, which can be used to explain decision-making behavior and problem solving strategy selection. Anderson (1990) proposes that the best method for analyzing human cognitive behavior lies in the analysis of the task rather than in attempting to analyze the methods used by the human to solve the problem. Crucial to the framework is the concept of optimality, in which a rational person behaves in such a way as to maximize the utility (in a particular situation) given the various constraints operating on the cognitive processes underlying their decision behavior. In this framework, the function of human memory is to retrieve knowledge that would allow the user to perform his task. The optimization process maximizes the expected utility of the memory system by balancing the cost of memory search against an expected gain of retrieving the desired item for the task. An example of cost of memory search is time. In Anderson's rational analysis framework, the human memory system would search a memory structure until the probability of getting the desired memory item (the expected gain) is lower than the cost of further search (i.e., when the expected utility becomes negative).

## 2.2 Limitations of Information Processing Psychology

One can go along with the idea that "thinking is information processing" in the sense that all neural activity can be regarded as processing. The stages of problem solving outlined above are valuable in many cases, but they might be too mechanical and too rational, assuming obedient problems solvers in isolated situations. It can fit well-defined tasks without interface interference, but it might be too limited for our HCI context. We refer to the fact that in the context of HCI, we have to keep human nature and the way we behave (habits, motivational factors) into account, combined with the fact that interfaces can influence this. Examples are how we are distracted during computer use, the habits we have developed, and the way we think we can rely on the assistance of software. Regarding plan-based behavior, Miller, Galanter and Pribram (1960) stated that planning activity could be seen as the mental evaluation of moves before they are executed, and behavior can be seen as the execution of pre-computed mental plans. Rational analysis, just as IPP (Newell and Simon, 1972) fits well with this view, and is valuable in many cases, but as said, we believe it might be too rational for our context. There is criticism concerning the way in which planning and action *in the real world* are interleaved and on the continuing reliance on gathering information from the world to instigate planning. Todd and Gigerenzer (2000) stated that humans make inferences about their world with limited time, knowledge, and computational power. This is often in contrast with models of rational inference that view the mind as if it would have very strong inclination to reason with unlimited knowledge, and without time restrictions to make decisions. They see such visions of rationality as conflicting with reality. IPP leaves relatively little space for human tendencies that can influence behavior, such as "having the feeling the work is partly done for the user" or "being directed towards a solution", and the influence this can have on the way people plan and execute their actions.

In this context, it is worthwhile to acknowledge that people do not always follow planned predefined processes. People are also influenced by their environment such as social elements, instinctive factors, values, and context. In the context of HCI, we have to keep human nature and the way we behave in account, combined with the fact that interfaces (the environment, the secondary task) can influence this. The almost purely cognitive angle on problem solving that rational analysis and IPP share has been criticized in favor of other accounts of behavior. One line of thought, which sees action as responsive to environmental circumstances is called *situated action* (Suchman, 1987). Important was the observation that when solving problems, people do not always plan complete sequences of actions before acting on the world. If this is the case, which aspects of a task encourage users to plan? An extreme example of findings that go against the notion of behavior as a purely planned activity is the earlier mentioned research by Mayes, Draper, McGregor and Oatley (1988). They showed that regular users of a Macintosh exhibited very poor recall for menu headers and menu labels, as if they never paid attention to where commands are. They concluded that people perhaps often decide what to do on the basis of information flow, and embark on activities expecting to find out how to succeed as they go, without a plan. This is reminiscent of Externalization as we use the term, and fits with display-based behavior. The authors see this kind of activity as common in our lives, and acting as the basis of how we interact with modern interfaces that minimize memory load: user behavior is guided from moment to moment by what users take to be clues. This approach is interesting because it suggests that the behavior of users of this kind of menu systems could be driven by the appearance of the display and the way it reacts to actions from the user rather than learned (and remembered) representations of the procedures.

## 2.3 Cognitive Load Theory

In computer interfaces, many signals and events have to be processed and paid attention to. To master the functioning of the application itself (the secondary task, see section 1.4), one can think of interface related items and events that play a role in the process, and consequently have their influence on thinking processes. When we use our applications, not all processing is actually geared towards appropriate processing as the designers of these systems intended (depending on the task of course). No matter whether the designer of the application desired it or not, if something starts blinking on the screen, a user will notice it and most likely look at it, and will (at least subconsciously) try to make sense of it. This is also processing, and regardless of whether this must be considered "thinking" or not, this is certainly *not* always serving the primary task.

Since the domain of this research is information from computer interfaces in combination with problem solving, and the focus is on how users can be provoked into certain *kinds* of thinking activity, we will adopt Cognitive Load Theory (CLT) as an addition to the information processing account. CLT has become a fundamental theory used to describe cognitive processes in learning with new technologies, such as multimedia environments. It has its roots in information processing theory, and was also developed during the study of problem solving (Sweller, 1988). The general goal of CLT is to develop instructional guidelines that enable learners to use their cognitive capacity as effectively as possible. As Kirschner (2002) states, CLT can provide guidelines to assist in the presentation of information in a manner that encourages learner activities that optimize intellectual performance. CLT is specifically concerned with the limitations of Working Memory, and the ways one can promote learning and the construction of schemata. A schema is a cluster of knowledge related to a problem type. It contains information about the typical problem goal, constraints, and solution procedures useful for that type of problem (Gick & Holyoak, 1983). CLT has the basic premise that learning is hindered if the instructional material overwhelms learners' cognitive resources. See figure 2.1 for a schematic overview. The theory differentiates between Intrinsic, Germane and Extraneous Cognitive load (Sweller, Van Merriënboer & Paas, 1998). Intrinsic Cognitive Load (ICL) is the load imposed by the task and refers to the complexity of a concept itself. The numbers of interacting elements and their relationships have to be kept active in Working Memory and this reflects the difficulty of learning (relational complexity, see Figure 2.1). Extraneous Cognitive Load (ECL) results from the design of the instructional material, and refers to the *unnecessary* load caused by inefficient instructional design. Germane Cognitive Load (GCL) refers to the degree of effort involved in the processing, construction and automation of schemas. GCL is the load that is *effective* for learning and is also associated with motivation and interest. Regarding GCL and motivation, Paas, Renkl and Sweller (2003) state that "*Note that increases in effort or motivation can increase the cognitive resources devoted to the task. If relevant to schema acquisition and automation, such an increase also constitutes an increase in Germane Cognitive load*" (p. 2). For some more discussion on CLT and motivation, see Paas, Tuovinen, Van Merriënboer and Darabi (2005). ICL is seen as unchangeable from the outside, whereas designers are thought to be able to influence ECL and GCL load by *triggering* learner activity via instructional designs. Central to CLT is the notion that there are different kinds of cognitive load, and that not all types of cognitive load hinder performance, it depends on the kind of cognitive load.

CLT has specifically focused on the representation of information, in particular with respect to instruction and learning material. Nevertheless, we believe CLT does not need to be limited to learning and instruction alone, but can also be applied in HCI. While this research is not focusing on instruction per se, we will adopt the assumptions of CLT, which also state that a good portion of WM should be devoted to effortful cognitive processes. Several researchers have proposed to adopt CLT in HCI research, e.g. in information retrieval, which can also be regarded as a problem solving process (Back & Oppenheim, 2001). One could associate Intrinsic Cognitive Load (task difficulty) with query difficulty and Extraneous Cognitive Load (task presentation) with interactive properties of the interface. The latter fits with the division between primary and secondary task, as we elaborated upon in section 1.4. Feinberg and Murphy (2000) applied CLT to the design of web-based instruction, and concluded that CLT provides general design principles to reduce Extraneous Cognitive Load, and that it is a good thing to develop graphical user interfaces and multimedia formats in consideration of cognitive load principles to effectively enhance web-based instruction. Also from a multimedia context, it has been stated that the demands by interactive media may exceed the capacity of the cognitive system and prevent learning (Mayer & Moreno, 2003). CLT also assumes a limited working memory connected to an unlimited long-term memory, and a coordinative component, the central executive.

Through the years, augmentations to Cognitive Load Theory have been proposed. For example, Rikers, Van Gerven and Schmidt (2004) mention that the theory in its original form could not account for the importance of the role of learners' expertise (in Figure 2.1, this is included). Gerjets and Scheiter (2003) propose an extension of the theory based on the assumption that instructional design is not *directly* determining cognitive load, but that learning activities (in terms of goals and strategies) are important moderators between instructional design and cognitive load. Gerjets and Hesse (2004) state that these broad principles by themselves might not be sufficient to ensure that effective learning activities will occur, and stated that learning activities might be influenced by other factors, such as learners' preferences, styles, attitudes, epistemological beliefs and prior knowledge. Still, CLT has proven to be a valuable framework for research, and for simplicity purposes, we will stick with the basic assumptions of CLT (see figure below) because they suffice for the issues we investigate.

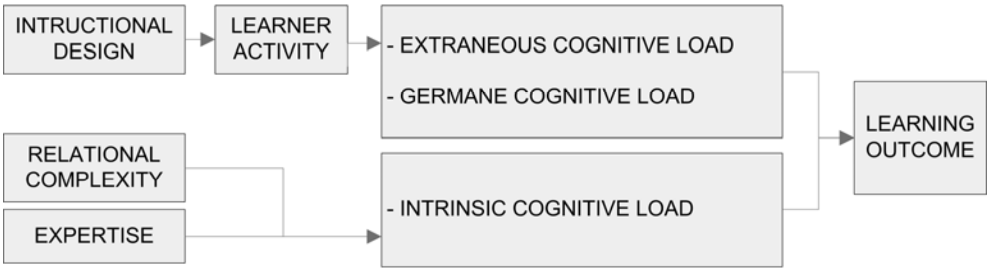



Figure 2.1 – Assumptions of Cognitive load theory (based on Gerjets & Scheiter, 2003)

## 2.4 Distributed Cognition and External Cognition

Several authors have used the term *distributed* in connection with cognition, acknowledging that cognition can be regarded as taking place in a distributed manner. Norman (1988) pointed out that information embedded in technological artifacts is equally important as the knowledge residing in the mind of the individual who used the artifact. He argued that well-designed artifacts can reduce the need for users to remember large amounts of information, while badly designed artifacts increase the knowledge demands made on the user. There are various ways one can look at having information "in the head" and "in the world", and these two concepts correspond with the two main scientific disciplines which research on the distributed nature of cognition combines consists of, i.e. cognitive psychology and anthropology. Theories of distributed cognition represent the claim that cognitive processes are generally best understood as situated in, and distributed across concrete socio-technical contexts. The term *distributed cognition* as such was coined by Ed Hutchins at the University of California, San Diego, and refers to the idea that cognitive activity is distributed across internal human minds and external cognitive artifacts (Hutchins, 1995). Hutchins' emphasis has been on team work and how cognition can be distributed across a group of individuals, but the approach can also be used for individual problem solving, where the emphasis is on how cognitive components of a problem are distributed between a single human problem solver and problem solving artifacts. A general principle of the distributed cognition approach for cognitive systems is that people off-load cognitive effort to the environment whenever practical (Hollan, Hutchins & Kirsch, 2000), and that artifacts and representations should be designed to maximize the potential for such off-loading (see also Zhang & Norman, 1994). The components of a distributed cognitive system are internal and external representations. Internal representations are the knowledge and structure in an individuals mind, and external representations are the knowledge structures in the external environment (Zhang, 1993, 1997; Zhang & Norman, 1994). Complex information processing tasks require the processing of information distributed across internal minds and the external artifacts. An example is a multiplication task using paper and pencil; internal representations are the meanings of symbols and procedures that have to be retrieved from memory; the external representations are the numbers written on the paper. The information from the internal and external representation has to be processed in a dynamic manner.

The term *external cognition* was coined by Scaife and Rogers (1996), and also refers to the interaction between internal and external representations (the latter being external memory) when performing cognitive tasks (e.g. learning). Scaife and Rogers provide a useful overview on the use of graphical representations in educational environments, and outline a theoretical framework. They conceptualize how different kinds of graphical representations (e.g. diagrams, animation, and multimedia) are used during cognitive activities. The framework presents a set of core properties, of which the central one is computational offloading (the extent to which different external representations reduce the amount of cognitive effort required for cognitive activities).

## 2.5 Related research

Above we introduced some frameworks and terminology to look at our research phenomena. We will here present more research that is relevant borrowing from these frameworks. Broadly, we will divide the research we mention in research concerning alterations in the representation vs. alterations in the interaction-controls.

### 2.5.1 Altering the representation: Computational offloading

The notion of computational offloading, concerning task and software supporting the ability to rely on display features, supports the idea that novice users should not need extensive knowledge (internal representations) of the interface, in order to use its basic features. We agree with this, as far as it concerns the type of interactions mentioned before, such as using ATM's, ticket dispensers, common off-the-shelf software, etcetera. Altering representations by externalizing certain information could have several advantages to the problem solver, for example the reduction of load on working memory. It can provide constraints on the range of potential user actions, which in turn reduces the size of the problem space (Zhang, 1997; Zhang & Norman, 1994).

Quite a large body of research concerning problem solving is done in the context of external displays (Larkin, 1989). Regarding external representations, Kotovsky, Hayes and Simon (1985) studied different versions of the Tower of Hanoi problem. They found that learning the rules of the legal operators were a major source of difficulty and that the versions with more difficult rules took longer to solve. Rules consistent with real-world knowledge and the presence of a physical, external representation aided problem solving by reducing the working memory load. Intensive training of the rules, even in isolation from any problem setting or knowledge, significantly decreased the difficulty of problems. This indicates that external representations of the goal and rules can improve performance. They state that working memory capacity is easily overloaded by the memory requirements of unfamiliar problem rules. This overload prevents even minimal amounts of necessary planning from occurring.

Kotovsky and Simon (1990) obtained more evidence that the type of information available to a problem solver and the problem representation influence problem difficulty. In an experiment, subjects solved different computerized isomorphs of the Chinese Ring Puzzle. The three versions they used were called “no-info” (no aid at all), “lo-info” (subjects were informed of legal moves) and “all-info” (legal moves are displayed, and the resulting state of each potential move). Their measures were the number of illegal moves, total number of moves, and time needed. It showed that subjects in the lo-info version performed better than the no-info version. The lo-info condition informed participants of which moves were legal and which moves were not legal (as in the destination feedback example, see below Figure 1.6), and this proved to be helpful. However, the all-info version scored worse than the other two versions, which makes interpretation more difficult. The all-info version was thought to confuse subjects. The conclusion was that giving additional information about move legality had positive effects only if the information was understandable and not confusing. Also in the studies of diagrammatic problem solving, Larkin and Simon (1987) and Larkin (1989), for example, also argue that diagrammatic representations support the recognition of features and the directly making of inferences.

Zhang and Norman (1994) and Zhang (1997) focused on the nature of external representations and the interactions among internal and external representations. They showed that external objects are not simply peripheral aids, but that they provide a different form of representation. According to this framework, the appropriate unit for cognitive analysis is not the human mind in interaction with an information-rich environment, but rather a work-system, potentially including multiple minds and artifacts. Zhang and Norman (1994) studied the effects of different distributions of information across internal and external representations on problem solving behavior in the Tower of Hanoi task. The problem has three rules, and they can be implemented in either internal or external representations. Three isomorphs of the problem were presented, one where all the rules were internal because they all had to be memorized; the second where 2 rules were internal and 1 external, and the third version where only 1 rule was internal and 2 were external. An experiment was carried out with these three isomorphs, and solution times, solution steps, and errors were measured. The problem space had a solution that consisted of seven moves. The results showed that more externalized rules resulted in better performance (the more externalized, the less time and moves needed and the lower the error rate). Based on these results, the authors made the following statements about external representations:

1. External representations provide information that can be directly perceived and used without being interpreted and formulated explicitly.
2. They can anchor cognitive behavior (physical structures in external representations constrain the range of possible cognitive actions in the sense that some actions are allowed and others prohibited).
3. They change the nature of tasks: tasks with and without external representations are completely different tasks from a task performer's point of view, even if the abstract structures of the tasks are the same.

### 2.5.2 Altering the controls: Increasing operator cost

While the research mentioned above concerns how representations can influence performance, there are also experiments showing that variations in the controls of interface operators can have an influence on planning and performance. In order to predict which characteristics encourage planning, O'Hara and Payne (1998) adopted the framework of rational analysis. From a rational analysis of problem solving they examined the assumption that increasing the cost of an operator increases the level of planning during problem-solving, and introduced the notion of *operator implementation cost*. By operator implementation cost is meant the fact that different operators, such as typing a command, or making a move with the mouse have different "costs". This cost consists of, for example, the amount of time, physical and mental effort associated with applying an operator (such as the time and physical effort associated with typing a command string to execute the command).

O'Hara and Payne (1998) did an interesting series of four experiments regarding operator implementation cost in command-based interfaces, in which a high-cost and a low-cost interface to a problem were compared. In the first experiment, using a task called the 8-puzzle, the operator cost was manipulated by varying the command-length to execute an action (shorter vs. longer). The results showed that the low-cost subjects performed faster, but this was logical regarding the longer sequences they had to type to execute commands (longer commands cost more time). The move analyses, however, provided evidence that is consistent with the idea that high-cost subjects work more plan-based: the high-cost subjects made fewer moves to reach the solution than the low-cost subjects did. The second experiment used the same material, but now verbal protocols were analyzed. The analysis of the verbal protocols and move sequences revealed behavior patterns that are consistent with the findings from the first experiment regarding the relationship between operator implementation cost and planning. In the third and fourth experiment, the influence of "doing a task training phase in one interface (training)" on "performance on a same domain task (transfer)" was studied. It showed that over the training trials the subjects using the high cost interface performed more efficiently in terms of solution lengths, replicating the finding of the first experiment. However, what is more interesting, and supports the basic cost-benefit framework, was the performance on the transfer trials (in which all subjects used the same interface). Subjects who were trained with the high cost interface performed better in terms of time and number of moves to solution. This strengthened the authors in the assumption that more reflective plan-based strategies by high-cost subjects led to better learning and transfer. They concluded that provoking planning helped to people to become more aware of the underlying structures of the problem rather than simply memorizing solution paths.

O'Hara and Payne (1999) conducted another series of three experiments in which through simple (but again different) user interface manipulations the cost of operations was varied. The first experiment uses the slide-jump puzzle, but this time they looked at the effects of imposing an implementation cost only on the undo operator task interface. The system imposed a high implementation cost on undoing incorrect moves. In line with the predictions of the cost-benefit model, this experiment again showed that efficiency (solution length, at least initially) was better for the high-cost interface. The second experiment investigates the influence of a system lockout delay after every operator application, on problem solving with the 8-puzzle. Lockout is the interval between system feedback and the point at which the system is ready for the next input. As expected, the longer the imposed lockout time, the shorter solution lengths were. Solution lengths can be seen as an indicator of the amount of search in the external representation of the problem space. The authors concluded that with a higher lockout time, the subject's search within the distributed problem space was done less by the strategy of manipulating the external task representation (acting) and more by the strategy of traversing the internal representation of the problem space (planning). The third and last experiment uses a manipulation of lockout time and attempts to replicate the effects within the domain of an office type task. In their experiment, subjects had to prepare letters that had to be sent to visitors of a conference, by copying the information and pasting it into the appropriate letter. Again, subjects that worked with the high time cost condition completed the task in fewer moves. With the high implementation cost, the benefits of investing more cognitive effort into further planning before acting were greater than in the low time cost condition, leading to less acting in terms of shorter solution lengths in the high lockout group. Generally, behavior in the high-cost conditions was characterized by shorter solution lengths and greater thought time per move. In the low-cost conditions, subjects showed more of a “trial and error” approach with low thinking time per move and longer solution lengths.

In an earlier research paper, after conducting comparable experiments, O’Hara (1994), concluded that the mechanisms of behavior are neither exclusively situated nor plan-based. Rather, the “playfulness” or “situatedness” of behavior is dependent on a number of factors, one of them being the implementation of the cost of an operator. He stated that:

*“it is not always the case that all aspects of the interface should be designed to be as easy to use as possible. For example, it may be desirable to increase interface cost in instances where one wishes to reduce the number errors or when there are limited financial and physical resources available to allow the trial and error type approach of the low cost interface users. In addition, there may be certain points within a particular task where thoughtless execution of an operator may have a detrimental outcome, thus at these points, it may be good design to increase operator cost in order to avoid these unwanted consequences” (p. 106).*

Concerning the latter, Svendsen (1991) conducted a study that compared the effects of a direct-manipulation and a command-line interface on a user’s ability to solve the Tower of Hanoi problem. Subjects who used the command-based interface were more efficient at solving the problem than those using the direct-manipulation interface, in that they made the least number of errors and reached the criterion in fewer trials. An explanation was that the different interface styles induce learning in one of two different types of learning mode: “Unselective-mode” (U-mode) or “Selective-mode” (S-mode). U-mode learning refers to "implicit" learning, and is thought to take place outside working memory. S-mode learning is comparable with "explicit", and is effortful and involves rule-generation (Hayes & Broadbent, 1988). Perhaps the feedback in the direct manipulation interface provokes U-mode learning, whereas command-based interface induces S-mode learning. However, the different effects of the two interfaces might be also explained by their different operator implementation costs. In the direct manipulation condition, performing an action is “cheaper”, in terms of effort, than in the command line based interface where commands have to be typed with the keyboard. Because actions are more “expensive”, users of the command-based interface are prompted to think harder, making correspondingly fewer errors and learning more efficiently.

In the same vein in the context of cost and benefit, Trudel and Payne (1995, 1996) compared exploratory learning of the functions of a digital watch with perhaps even more dramatical results. They imposed a keystroke limit on the exploration (hi-cost condition) or no keystroke limit (low-cost condition). It showed that high-cost subjects were more efficient at reaching a test goal, using significantly fewer moves, and also learned more. The keystroke limit made interactions a more valuable resource, and thus encouraged subjects to pay more attention to, and think harder about the results of their actions.

Again, a different approach is by Gray and Fu (2001), who studied users when they learned to program a simulated VCR. There were three conditions: “memorizing the procedure in advance", "inspect the next step looking at a help box with high-cost" and "inspect the next step by looking at a help box with low-cost". The "memorizing the procedure in advance" condition had the best performance (time and errors). Then, within the conditions that could consult help subjects in the high-cost condition often preferred to not consult the help at all, but relied on their imperfect memory, thus making more errors than the help condition that consulted the help window simply by looking at it. It showed that people do not always choose to offload memory demands to the environment, but weigh the relative costs and benefits of knowledge-in-the-head against knowledge-in-the-world.

### 2.5.3 Our approach: Destination feedback by varying the extent in which operators externalize information

We are on the one hand inspired by the research mentioned above, demonstrating that the cost of an operator can be of influence on cognitive strategies (O'Hara & Payne, 1998, 1999; Svendsen, 1991; Trudel and Payne, 1995, 1996), and on the other hand by research advocating advantages of Externalization (Zhang & Norman, 1994; Kotovsky & Simon, 1994; Larkin, 1998).

We do not intend to replicate or falsify results such as the ones of Zhang (1997), Zhang & Norman (1994) or Kotovsky and Simon (1990), but we focus on a slightly different aspect of Externalization and Internalization. We agree with the mentioned authors that external representations are not simply inputs and stimuli to the internal mind, and they are much more than memory aids. For many tasks, external representations are so intrinsic to the tasks that they guide, constrain, and even determine the pattern of cognitive behavior and the way the mind functions. It is the second remark (in section 2.5.1) that Zhang concluded with, that we dare to challenge. It is the beneficial consequences of the second statement, which we question: showing what is allowed and what is not, and the way in which this anchors cognitive behavior.

If one thinks only in general terms about the opposition between "making it easier" (alter representation for better performance) or "making it harder" (alter operator/controls for better performance) it could seem that the findings of mentioned authors are contradicting. However, if one scrutinizes more, it becomes clear that the two currents of research (alter representation vs. alter operator/controls) are not contradicting each other on the same dimension. We want to connect these currents and position our angle on the matter as being *between* these two. We will not alter the representation or the operators per se, but study a combination of these two; we will vary the external *representation* via the *operator*. In how we implement Externalization, the controls in the interface themselves become the conveyor of externalized information. This type of Externalization is called destination feedback, and in fact is what is implemented in the chess example of Figure 1.5. It does not actually change the problem space itself, but emphasizes directions that are possible, which can guide decisions. Destination feedback has been widely advocated by authorities in providing guidelines (Sun, 2001; Apple, 1992) and is default in modern computer interfaces. We will vary Externalization in the way the horse can be clicked in our chess example; by showing what a certain object or objects can do *at this moment*. This could remind one of the term *affordance* as mentioned by Norman (1988), but it is not the same. Affordances show the general properties, how an object may be interacted with in a general way (in the chess analogy: what can be done with a horse in general).

Destination feedback as we will implement it externalizes the *consequences and enforcement* of certain rules. When a user drags an item from its place to a destination, the application will provide feedback that indicates whether it will accept that item. What happens in the chess example can even be thought of as “feed forward” because it goes one step further since it does not only communicate that something *is* an operator or what it generally does. This could seem and feel like guidance or assistance that makes the task easier because fewer inferences and computations have to performed by the user (as in Figure 1.6). This is different from for example Zhang’s (1997) Tower of Hanoi (TOH) experiments. In the different versions of TOH, he varied how many rules were continuously externalized (visible to the user, as in affordances), in the puzzle as a whole. There was no dynamic interactivity as in destination feedback, the Externalization of objects was static and constant and not dependent on where one is in the problem space. Also in the context of research in which operator cost is varied, we do not make operators more expensive in the sense that delay occurs, or that lockout takes place. Instead, we vary the “friendliness” of the controls, interface elements that do, or do not convey information that can be perceived as assistance or guidance.

In essence, our manipulations come down to variations of the amount of interactivity an interface with its controls has, and we are interested in how this influences users to reason and make inferences. In the context of the terms primary and secondary tasks, if we see the task, the task that has to be performed as the primary task and the interface as the secondary task, we will focus on how the secondary task and its properties can, or cannot provoke effort, contemplation and planning for the primary task.

## 2.6 Cognitive Load Theory connected to our research

CLT can be a promising theory for our research issues, since the notion of different kinds of cognitive load can be seen as corresponding with the dual nature of computer-based tasks (primary and secondary) and might be better suited than IPP *alone* to accommodate our research. The way CLT distinguishes *effective cognitive load* (GCL) and *ineffective cognitive load* (ECL) could be a better way to look at the way users tackle problem solving with software that has its own characteristics and behavior. We question whether assistance implemented by Externalization to relieve cognitive load really *is* beneficial in certain situations. It might relieve CL in a certain way, but it could be CL of the wrong kind that is lowered, namely Germane CL, which must be high for good performance.

Also other researchers have used the concepts of CLT to describe this issue: assistance that could be counter effective. Scott and Schwartz (2007) investigated the role of navigational displays with informational websites of which the content had to be learned and comprehended. One way to reduce cognitive load in hypermedia navigation is to provide learners with a navigational map. They are meant to reduce the disorientation that learners experience, by making the navigational structure of the site explicit. Scientific results so far are rather conflicting, some researchers found beneficial effects (Chen & Rada, 1996; Simpson & McKnight, 1990; Utting &Yankelovich, 1989) but others, such as Stanton, Taylor and Tweedie (1992) found that site maps resulted in poorer performance, less use of the system, lower control, and poorer development of mental maps compared to control groups who did not use site maps. Delving deeper into this, Scott and Schwartz

(2007) ponder that it is not clear at all why navigational maps should reduce CL during navigation in the first place. Although learners might be informed of their location, they questioned the underlying cognitive activities caused by the maps. Maps offer more, not less information, which according to them could increase, not reduce CL. They questioned the presumed reduction of cognitive load caused by the map stimulating deeper comprehension of the information contained *within* the site, which is in the end the goal: learning information. Although the mentioned research concerned websites, it can be compared to computer-based problem solving. One has an initial state, a goal to accomplish and one mentally has to visualize the information space one is in, which can be seen as analogue to visualization of a problem space in problem solving. Also Williams and Noyes (2007) used CLT in research concerning external representations and the encouragement of higher order thinking strategies in which they showed that the design of the interface had direct implications on Extraneous Cognitive Load. Oviatt, Arthur and Cohen (2006) also used CLT researching new interface directions for math education. They stressed the importance of focusing attention on the math problems rather than on gratuitous features that may entertain, but also distract (i.e. minimizing complexity).

Regarding Germane Cognitive Load and the importance of it being facilitated, there is interesting research by Schnotz and Rasch (2005) which could support our assumptions. They focused on the conditions under which animated pictures enhance comprehension and learning. Animations, just as external displays, are thought to reduce cognitive load of tasks that could otherwise only be solved with high mental effort, and thus can have a facilitating function. However, they found that this facilitating function was not always beneficial for learning, because learners were prevented from performing relevant cognitive processes on their own. The external support made processing unnecessarily easy and, thus, students invest less cognitive effort in learning. From the perspective of Cognitive Load Theory, animation can unnecessarily reduce Germane Cognitive Load associated with deeper meaningful cognitive processing (Sweller, 1999; Sweller et al., 1998; Van Oostendorp & Beijersbergen, 2007). This is explained as a result of an (unintended) decrease of Germane Cognitive Load, because available mental capacity is left unused for the process of learning.

Also Guttormsen, Schar, Schluep, Schierz and Krueger (2000) recognized the importance of other interface design standards than *only* those suggested by common usability guidelines, and investigated the effect of the computer interface on learning performance. They employed a theory of explicit and implicit learning. Explicit learning is characterized by rational, selective, and conscious attention, and is more demanding for the cognitive system. Implicit learning on the other hand is similar to trial and error learning. With several problem solving experiments they showed that interaction tools that imply a *high* cognitive load on students when learning (e.g. having to *type* commands to interact) can induce an explicit learning style. Interaction tools implying little cognitive load on the users (e.g. direct manipulation with graphical objects), tend to induce a more implicit, trial and error learning mode. The type of task they used were for example computer aided drawing tasks, interactive exploration of Tower of Hanoi, learning rules for illumination ergonomics by interacting with a simulation or information search by navigation in a hypertext environment. They concluded that the user-interface can play a major role in the success of learning by supporting or inhibiting certain strategies. They propose to extend usability guidelines with extra cognitive factors, because usability (including user satisfaction) should not be optimized at the cost of learning performance.

We map the assumptions of CLT to our concepts of Internalization and Externalization. In a situation as in Figure 1.4 where the primary and secondary task are as separated as possible, this would result in the following (Figure 2.2):

1. **Intrinsic CL** depends on the complexity of on the primary task itself, and is seen as unchangeable
2. **Extraneous CL** (bad) is influenced by the functioning and behavior of the secondary task (interface)
3. **Germane CL** (desirable) is also being influenced by the secondary task (interface)

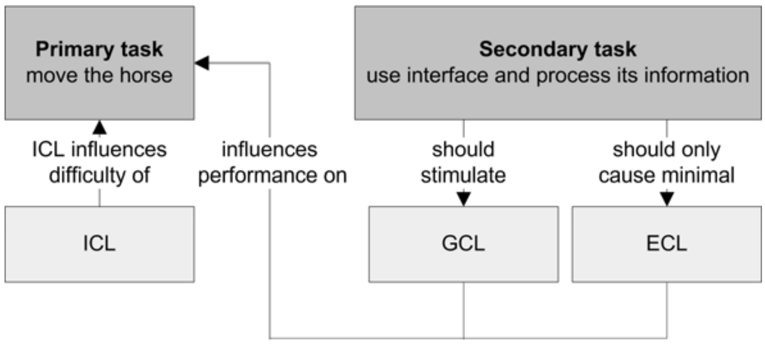



Figure 2.2 - CLT mapped to primary task and secondary task

We can work out more precisely what could happen during interaction when certain task relevant information is being externalized to the user meant to relieve cognitive load, as opposed to when users have to internalize all the information themselves. Below are the flows of events that could be caused by the secondary task (interface) events. Two scenarios are worked out in Figure 2.3, the path with the dashed line corresponds to scenario 1, and the path with the continuous line corresponds to scenario 2.

**Scenario 1. Secondary task with interfaces that externalize information**
Externalized task information makes recall of certain knowledge elements unnecessary, and is supposed to relieve WM, in the sense that certain mental activities are not necessary. The consequences can be twofold:

1. **Extraneous Cognitive Load (unnecessary load) is high: processing interface information takes resources away from the primary task.** The externalized information is meant to help, but still takes up attention and will *also* be perceived and processed, which can have distracting properties. In any case, it is *not* directed at proactive contemplation. Fewer resources are left for Germane Cognitive Load.
2. **Germane Cognitive Load will be lower than desired because of two reasons.**
   a. GLC has less resources available than desirable, because ECL is higher than necessary.
   b. Besides the above, the human tendency to feel that the task is (partly) being done by the system can already have an effect on GCL by itself. It can lead to less focus and proactivity and effort to solve the problem. Users are induced to plan *less,* and fewer resources will be devoted to problem solving and activities directed at

schema acquisition. Performance will be less efficient, and imprinting of strategies could be worse.

**Scenario 2. Secondary task with interfaces where information has to be internalized**
Internalized information must come from the user's memory. Cognitive load might be higher, but:

1. **Extraneous Cognitive Load (unnecessary load) will be low: no resources are taken away from the task**. Interference in the way of interface feedback of cues is absent, ECL is minimal. Users have to do everything themselves, nothing will give them the impression they are being guided or assisted.
2. **Germane Cognitive Load, devoted to mindful cognitive processing is allowed to be high.** ECL being low leaves a maximum of resources free for GCL, the load that is *effective* for problem- solving, and is also associated with motivation and interest. It concerns load which is associated with the degree of effort involved in the processing, construction and automation of schemas. The consequence of increased Germane Cognitive load is more plan-based behavior and consequently more elaborated strategies leading to better performance by users.

Below in Figure 2.3 we visualized the above scenarios in relation to the assumptions of CLT.

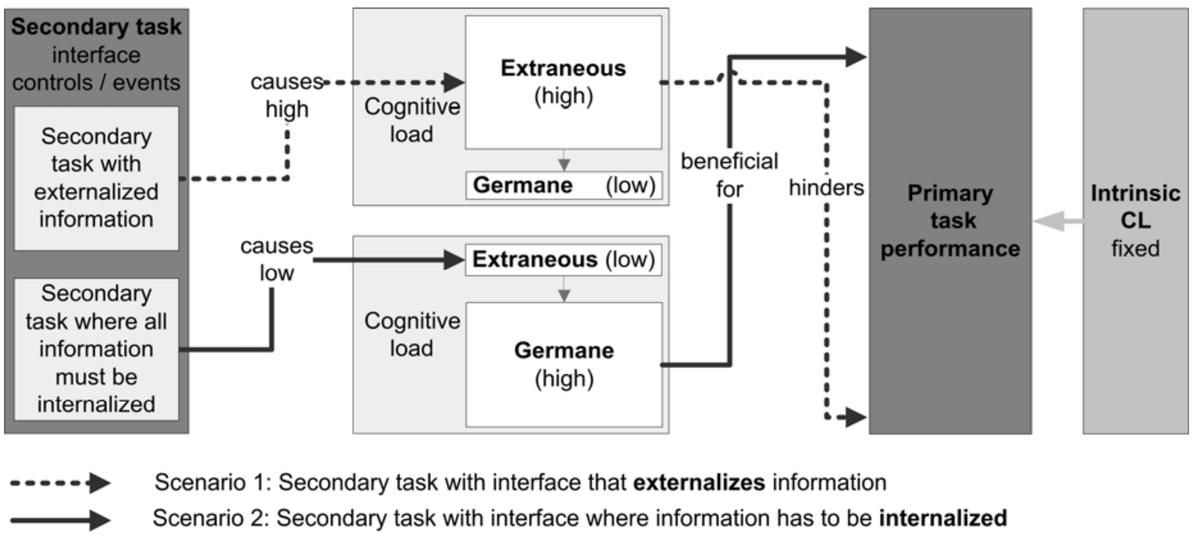



Figure 2.3 - CLT mapped to primary task and secondary task

We have now laid out our motivation, research scope, the issues it concerns, and the theoretical framework we position it in. The theories and research mentioned above are all valuable to take into account. On the one hand, we were triggered by findings that Externalization of information is seen as beneficial, and on the other hand, that making a task harder can be beneficial as well. We question whether the presumed facilitating qualities of computational offloading by for example Externalization of information really can be said to be beneficial. We ponder that cognitive load is not necessarily bad; it is the type of cognitive load that matters, as clearly has been shown by the experiments of O'Hara and Payne (1998, 1999). Germane Cognitive Load is the type of load that has to be facilitated, which corresponds with the planfulness mentioned by others which in turn facilitates strategy construction and mindful contemplation. The terminologies of cost-benefit trade-offs as mentioned in rational analysis and in the O'Hara and Payne research, the concept of situated action acknowledging environmental variables, the assumptions of the CLT approach and finally IPP will all be used in the remainder of this research. Using the strengths of various angles of viewing cognitive behavior (which sometimes have a certain overlap), with CLT being the central one, will aid in presenting an interaction framework as we announced in the final part of the introduction. The difference with the most common use of CLT is the idea that CL should be kept as low as possible and that this will improve performance.

It is clear from the mentioned findings regarding altering interface controls and cost of operators, that counterintuitive design decisions can have beneficial effects. These remarkable findings concerning making systems "harder" to use fit well with our assumption that the reduction of cognitive load and making interactions as smooth as possible is *not* desirable in all situations.

**Connection to the research questions**

Going back to our research questions (section 1.5), we will examine the influence of the interface on user behavior in solving problems requiring planning, be it positively or negatively. We investigate the conditions under which *externalizing* interface information by destination feedback influences users' plan-based behavior and strategy, and consequently performance in solving problems. Furthermore, we will investigate which interface style causes the best knowledge about the tasks, rules and solutions afterwards, which interface style leads to the best performance in a transfer situation, and which interface style provides better resistance against a severe task interruption.

Generally spoken, we assume that implementing Externalization in certain tasks leads to shallower processing, less proactive contemplation, less metacognitive activity, and consequently worse performance (less efficient solutions) on problem solving tasks. An interface where users have to internalize certain information by themselves without guidance might seem a more unfriendly system, but is expected to result in more efficient solutions, and better imprinting of knowledge in memory. We will start by gathering empirical evidence for our assumptions in the next chapter, in which destination feedback as explained in section 2.5.3 will be varied in the interface to a computerized isomorphic version of the well-known Missionaries & Cannibals problem.

# 3 Experiments with a puzzle-like problem solving task

In our experiments, we investigate the influence of interface presentation (with or without Externalization) in tasks requiring planning, on various determinants of performance. We aim at obtaining more insight into the role Externalization and Internalization play in provoking plan-based vs. display-based problem solving and the mental processes involved, in order to be able to contribute to a predictive theory of when and where to apply Externalization features (or not) in the interface. The focus in our experiments will be on deliberately varying task related information that is conveyed by the interface controls itself and thus varying the earlier mentioned separation between primary and secondary task by letting the interface operators become a part of the to be solved problem. In our experiments, users will have to solve problems in two conditions: using an Externalization interface style which provides information about the legality of performing actions, or an Internalization interface style where this is not done. In this chapter, in experiment 1 (which consists of 2 sessions) and experiment 2, the material that is used is based on a classic puzzle, the Missionaries & Cannibals (M&C) problem. In experiment 1 session A we will vary externalized rule-related task information in tasks where planning and deeper contemplation is required and investigate its effect on performance and knowledge. Experiment 1 session A will mainly address specific research question 1 (section 1.5.1) concerning how problem solving performance evolves over time, and how interface style influences this. In Experiment 1 session B takes place after a considerable delay. Besides the questions of session 1A, we investigate the influence of interface style on transfer of skill to another task (question 2), and whether the two interface styles have different effects on performance after a large delay (question 6). In experiment 2, the focus will be on the influence of instruction before starting the task. We will address question 5, which investigates whether instructing users to plan vs. not doing this influences behavior, and whether there is a difference between the two interface styles.

## 3.1 Material: The problem and its problem space

The Missionaries & Cannibals problem (Ernst & Newell, 1969) is a well-known logic puzzle, repeatedly used as an example of using abstract concepts to solve a problem. The most well known version of the problem consists of 3 missionaries and 3 cannibals that are standing together on a riverbank, graphically represented in Figure 3.1. The mission is to find a schedule for ferrying all the cannibals and all the missionaries safely to the other riverbank by boat (there is no other way). However, there are several constraints and rules:

1. Only 2 entities fit in the boat at the same time
2. The boat cannot cross the river empty; the minimum to sail is 1 entity
3. Cannibals can never outnumber missionaries at any of the two riverbanks, or the missionaries will be eaten (unless there are zero missionaries, then there are no missionaries to be eaten).

Figure 3.1 - A semantically rich version of the Missionaries & Cannibals problem

The choice of how the problem above is visualized most likely plays a role in the way the problem will be solved. The visualization above carries quite some semantic information in it and common sense becomes in a way part of the problem space. Certain information can be seen as continuously externalized already, such as the notion of who are the "bad" or "dangerous" ones. Cannibals are famous for eating people, and in Figure 3.1, rule 3 is continuously visually present. Rule 2 is graphically represented in the way that size-wise it is visible that the boat is full when 2 creatures are in it. Rule 1 also makes common sense because a rowing boat in real life cannot move empty either. This information is quite easy to learn and remember. In our experiment, we will use a variation on the problem that is devoid of all this directly available semantic knowledge represented to the user. This new more abstract problem will exist in two interface styles: an *Internalization* version where nothing is externalized and an *Externalization* version where the legality of certain moves with objects is externalized to the user (see section 3.2). We developed an isomorphic version of the M&C problem, which we call "Balls & Boxes" (B&B). It has more or less the same problem space as M&C, but is more abstract. We changed the following features (Figure 3.2):

- There are boxes instead of riverbanks, yellow and blue balls instead of missionaries and cannibals and a dish instead of a boat. By doing this, we avoid rules to be too easily learned at first, which makes it necessary that these rules really have to be remembered onward. It takes more profound imprinting in memory that "blue" are the bad ones, and "yellow" the good ones than recalling a fact that humans know: cannibals eat people and they are dangerous.
- The dish is quite big and does not visually give away that only 3 balls at a time can be transported. This has to be remembered by the user.
- We use 5 of each entity instead of 3. The maximum amount to sail then changes to 3 instead of 2. This does not alter the algorithm to solve it or the number of "crossings" in which the problem can be solved. It involves only more actions in terms of putting objects in the dish, not the number of crossing moves.

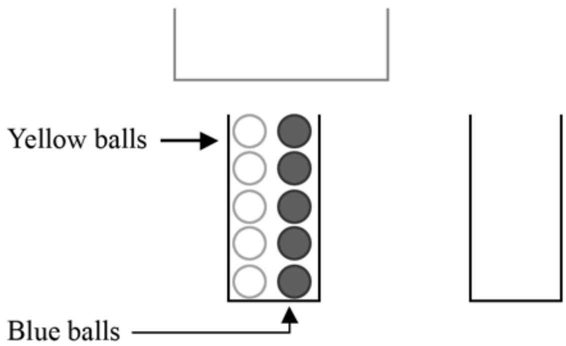



Figure 3.2 - The Balls & Boxes problem

The "new" rules of the Balls & Boxes puzzle are:

1. Only 3 balls can be transported with the dish
2. The dish cannot move empty; the minimum to move is 1 ball
3. Blue balls are never allowed to outnumber yellow balls at any side (only when there are zero yellow balls).

The application is made in such a way (see Figure 3.4 and Figure 3.5) that balls have to be moved up into the dish one by one by the user, and the action of moving the dish from side to the other also has to be initiated by the user. When the dish reaches the other side however, the balls will *automatically* drop out of the dish into the box below, so the action of dropping balls in the box after it has moved is performed by the program. In Figure 3.3 the problem space of the B&B problem is depicted, and it is visible how the goal state can be reached from the initial state. Starting from the initial state, the most economic solution consists of 11 dish-moves and 22 ball-moves, a total of 33 moves. Dish moves are the left-right movements of the dish; they are the transitions between the boxed problem states in Figure 3.3. Ball-moves are moves that put the balls into or out of the boxes before it moves (numbers are printed between the boxes in Figure 3.3, e.g. "02" indicating 0 yellow balls and 2 blue balls). The bottleneck consists of the states in the center of the problem space that have to be passed in exactly that sequence, they are mandatory to solve the problem. The boxes that are in gray color are *dead-end-states*, states that can be legally reached, but it is not very smart to go there, because they are like dead alleys: the only thing one can do after reaching them going back.

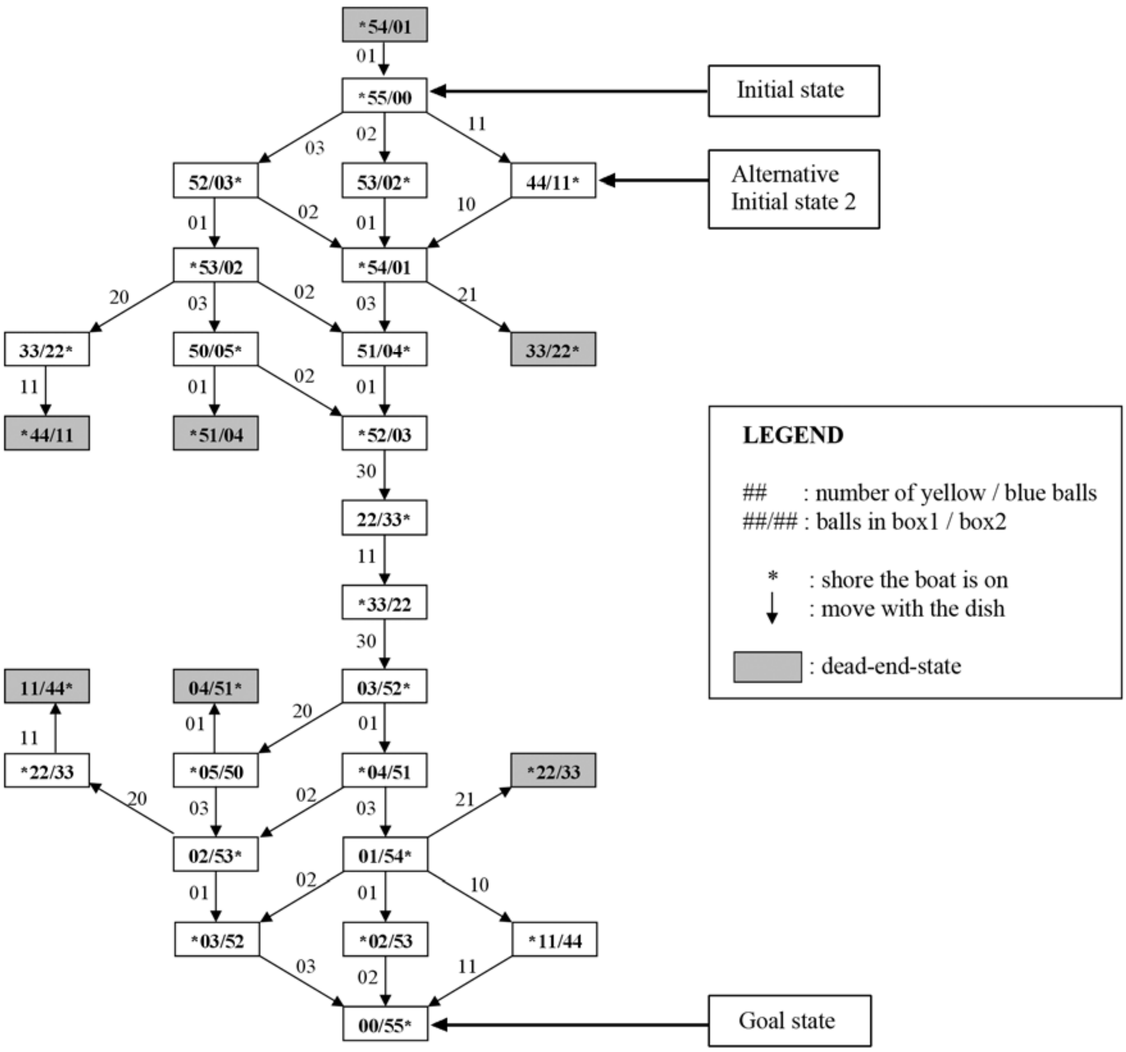



Figure 3.3 - The Problem space of Balls & Boxes

## 3.2 Balls & Boxes: implementing Externalization

Above we outlined the problem we will use in our experiments. A JAVA application of the Balls & Boxes problem (in two versions) was developed at Utrecht University by Sandor Spruit, which logged most of the users' moves and actions. The Externalization version is designed in such a way that the controls of the interface *itself* become part of the primary task. Solving the primary task (in essence a mathematical problem), and secondary task (the interface) will be intertwined to some degree. In the first experiment (which consists of two parts, session A and session B), the B&B problem will be used in 2 versions: an Internalization version and an Externalization version. By externalizing certain information onto the interface, certain rules and their implications are externalized to the problem solver and do not have to be inferred anymore. The controls of the application are arrows belonging to objects needed to perform moves in the problem space (Figure 3.4 and Figure 3.5). The arrows are colored (indicating they are clickable at this moment) when the *outcome* of the move would not violate any of the rules. Externalization is realized by presenting the rules explicitly on screen using visual cues. If clicking a control violates a rule, the arrow is grayed out, rendering it impossible to use. For example, in Figure 3.4 it is impossible to use the pink arrows to transport the dish, because there are no balls in the dish yet. In the Internalization condition, the rules have to be remembered by the user.

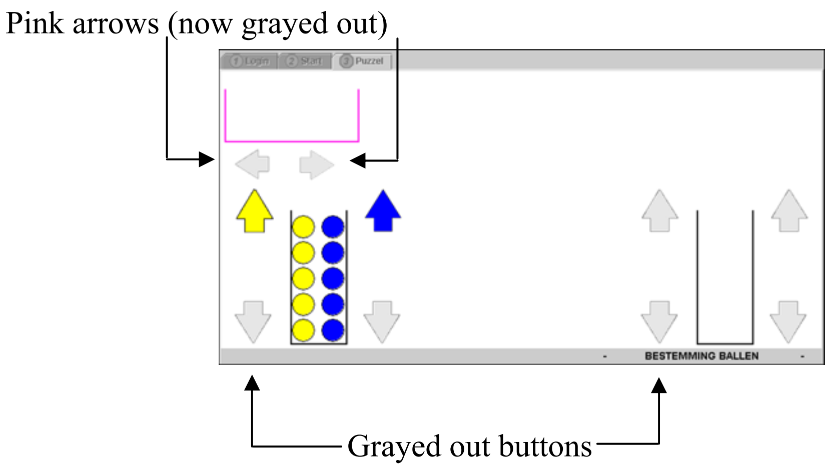



Figure 3.4 - Externalized version: information whether an action is possible (grayed out items)

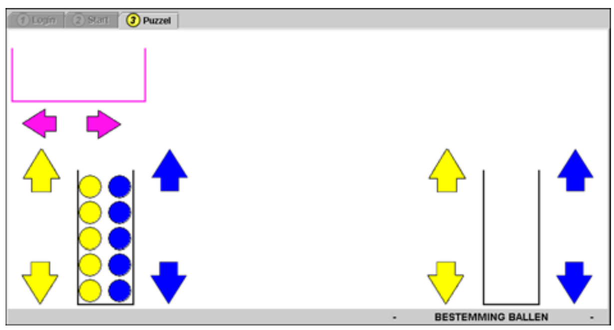
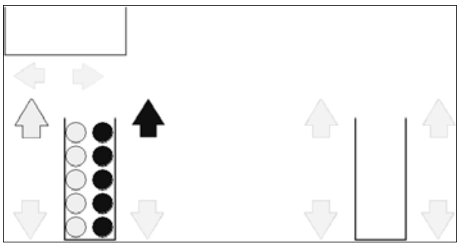



Figure 3.5 - Internalization version: no information whether an action is possible (all controls are clickable)

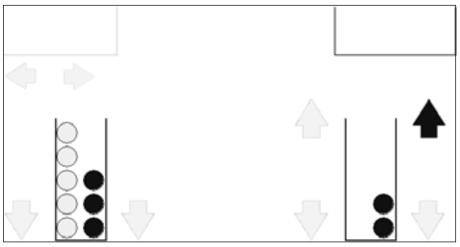
Regardless of the interface version, if no rules are violated, the following actions can be performed:

- Putting balls from the boxes up into the dish when the dish is above a box. This can be done with the up-arrows, the blue arrows for blue balls and the yellow arrows for yellow balls. When clicking on an arrow, one ball at a time will be moved up into the dish.
- Putting balls from the dish back down into a box (before moving the dish, when the dish is above the box). This can be done with the down-arrows, the blue arrow for blue balls and the yellow arrows for yellow balls. When clicking on an arrow, one ball at a time will be moved back down.
- Transporting the dish from the left side to the right side. This can be done by clicking the pink horizontal right-arrow that is below the dish. After the dish moves to the other side, the ball(s) will automatically drop down into the box.
- Transporting the dish from the right side to the left side. This can be done by clicking the pink horizontal "left"-arrow that is below the dish. After the dish moves to the other side, the ball(s) will automatically drop down into the box.

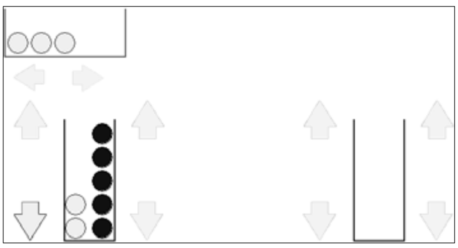
In this version of the application, it was also possible for users to consult the rules by clicking on a rules-tab. Below are several examples illustrating how the controls to do the task convey information about the rules (in the Externalization version). It is visually represented which moves are legal in any current state of the problem. On the right side of the examples one can read (in bold text) which rule is being externalized.



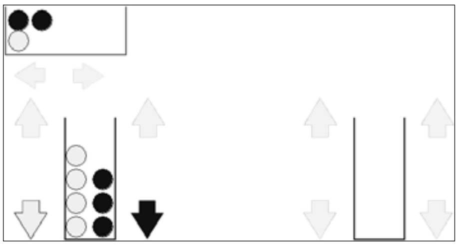
**Rule 2: The dish cannot move empty**
There are no any balls in the dish, so it cannot move: the two horizontal arrows below the dish are gray.
The only possible actions are to put yellow or blue balls into the dish, indicated by the arrows that have color. Moving balls out of the dish is impossible, since there are none in the dish.



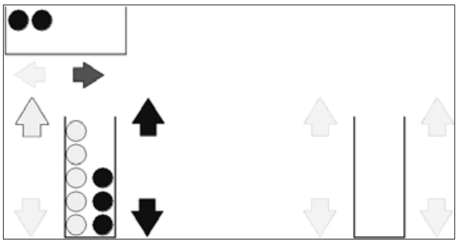
**Rule 2: The dish cannot move empty**
There are not any balls in the dish, so it cannot move: the two horizontal arrows below the dish are gray.
The only possible action is to put one or more balls from the right box in the dish (after this the dish can move).



**Rule 1: No more than 3 balls fit in the dish**
There are already 3 balls in the dish. The up-arrows are gray, telling the user that no more balls can be added to the dish.
**Rule 3: Blue balls cannot outnumber yellow balls**
If the dish moves, blue balls would outnumber the yellow ones on the <u>*left*</u>. This is illegal, so dish-arrows are gray.



**Rule 1: No more than 3 balls fit in the dish**
There are already 3 balls in the dish. The up-arrows are gray, telling the user that no more balls can be added to the dish.
**Rule 3: Blue balls cannot outnumber yellow balls**
*If* the dish moves, blue balls would outnumber the yellow ones on the *right*. This is illegal, so the dish-arrows are gray.



It is legal to move the dish to the right box: it would not create a situation in which a rule is violated. The dish-arrow is on (pink). It is legal to add a least one ball to the dish (yellow or blue). If one adds a yellow ball though, *then* transporting the dish becomes illegal: the pink arrow will turn gray.

Figure 3.6 - Several examples of externalized information telling users about rules and actions (yellow balls in gray, blue balls in black)

The Externalization of the rules causes that certain rules do not have to be remembered all the time and narrows the possibilities of user actions. However, it is important to realize that this visual aid by no means tells a user how to do it, it only informs a user of what is allowed given the situation.

## 3.3 Experiment 1 session A: Interface style (Internalization vs. Externalization)

Experiment 1 consists of two parts, session A and B. In experiment 1 session A, only the Balls & Boxes application will be used. Session B takes place after an 8 month delay and uses Balls & Boxes and in addition the above version of Missionaries & Cannibals, to see whether transfer of skill is taking place.

### 3.3.1 Design and subjects

Our design has one random-assignment between-subjects independent variable: Internalization vs. Externalization. Twenty-nine subjects of age 19-28 (mean age 21.8), experienced with PCs, were recruited at Utrecht University and were paid €5 for their participation. There were 14 and 15 subjects in the Internalization and Externalization condition respectively. The puzzle, a Java-applet, ran on a Pentium IV 2.0 GHz PC with a 17” monitor.

### 3.3.2 Measures

The application logged most of our dependent measures, which were:

- Number of puzzles subjects solved per phase, with a maximum of 3 per phase

- Time needed to solve the puzzles. We can only measure this in the trials that were actually solved
- Number of moves made. Regardless of whether the problem is solved or not, in the time subjects have they wander around the problem space, trying to solve the problem. We can have a look at the number of legal moves that were made while attempting to solve the problem in the time they had available. We expect the Externalization interface style to cause more actions, since we except more display-based and trial and error behavior
- Superfluous moves. In the problems that *were* solved, we can see how many extra user actions were performed, besides the optimal solution path
- Reaching dead-end-states. These are puzzle states that are off a direct solution path, indicating trial and error search (gray states in Figure 3.3). The less we expected plan-based problem solving to occur (Externalization) the more these states will be reached
- Knowledge test. After the puzzles we measured implicit and explicit knowledge of the rules and shortest-path solutions of the problem, using a paper questionnaire (see Appendix A)
- Attitudes: Likert-scale questions concerning perceived amount of planning, feeling lost during interaction and other such measures (see Appendix A)

### 3.3.3 Procedure

We informed the subjects of the course of the experiment, and gave them a brief explanation of the interface and a short demonstration. After that the subjects received an instruction that told them that all the balls have to be transported to the other side, but that there were some rules applying to the puzzle (which they could consult). Experiment 1 session A consisted of 9 B&B puzzle-trials divided into 3 phases. Slightly different initial states of the puzzle were used to avoid subjects simply repeating actions. The three tasks in phase 1 started with the initial state, and the two alternative initial states respectively (Figure 3.3). In phases 2 the 3 states were the same as in phase 1, only the playing direction was reversed from left-right to right-left. In phase 3 the initial states and playing direction were the same again as in phase 1. Between phase 2 and 3 all the subjects had a 10-minute pause from B&B and received a spatial ability task (see section 3.3.4) that served as a mental distraction. The maximum time for each trial was set at 7 minutes. In the third phase, subjects resumed solving three more B&B tasks, which were the same as in phase 1 (the playing direction was set to left-to-right again). After phase 3 subjects filled out a knowledge test (score 0-8) consisting of 4 multiple choice and 4 open questions with screenshots of puzzle situations. They had to judge and explain whether and why certain actions were possible (implicit knowledge), and recall the rules (explicit knowledge). Subjects also rated how clear the rules were for solving the problem.

### 3.3.4 Material to distract users mentally: the visual rotation task

One of our assumptions is that one interface style (Internalization) causes better and longer lasting imprinted task and rule knowledge than the other interface style (Externalization). In order to obtain evidence for this assumption, we will try to mimic a delay, in which subjects are doing something completely different, and then look at the ease of solving after this distraction. After solving a series of B&B puzzles subjects will probably have some routine to solve the puzzles. After a while, they will be told to stop doing this task and they will receive different task (they are not told that they have to resume more B&B tasks later on).

During these 10 minutes, the subjects will perform a task which can be expected to erase all B&B related routines from working memory. The task is a spatial ability task, in which visual rotation of figures has to be mentally performed (Figure 3.7). The test was provided by TNO – Human Factors Institute (Neerincx, Pemberton & Lindenberg, 1999).

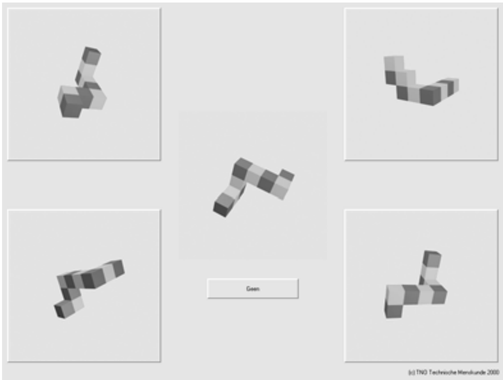


Figure 3.7 - Visual rotation task

Subjects are presented with screens displaying 5 geometrical shapes. The question to answer is which of the figures in the four corners is the same as the one in the center of the screen. Subjects have to choose between one of the four figures in the corners, or click on a button saying "none" if they think none is the same. The reason to use it here, as said is not our interest in spatial ability per se, but the fact that it is a highly taxing mental task that requires quite some mental resources. It will completely take away attention from the puzzles solved earlier.

### 3.3.5 Hypotheses for Experiment 1 session A

**H1: Externalization will *initially* yield better task performance than Internalization.** In the Internalization condition subjects do not have any task knowledge available yet. Externalization offers more interface cues; the guidance of indicating the legality of moves will aid the Externalization subjects in the beginning and result in better performance compared to the Internalization subjects.

**H2: Internalization yields better task performance later after pause with a distracting task.** After a while all the subjects will have learned to solve the puzzle. Internalization subjects had no guidance and had to acquire the solving skills by themselves. They will eventually store the rules in long-term memory more solidly, and have the needed information more readily available and perform better. This will especially be the case after the pause that erased working memory. Because of the guiding nature of the interface, Externalization subjects will plan and think less than the Internalization subjects, therefore work more on a trial and error basis and consequently display worse performance. We expect stabilization to occur between the two conditions in phase 2 and after that we expect Internalization subjects to perform better in the last phase.

**H3: Internalization yields better knowledge afterwards.** Not having externalized information available will motivate a subject to start planning on the basis of self-acquired rules. After the experiment, we expect the explicit knowledge of rules to be memorized better by Internalization subjects. There may also be a difference in implicit procedural knowledge, but we expect it to be smaller since both groups will have played the puzzle a similar number of times.

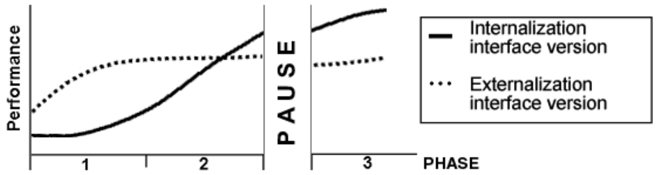



Figure 3.8 – Hypotheses experiment 1 session A

### 3.3.6 Results Experiment 1 session A

**Number of puzzle trials solved**

Subjects had to solve three puzzles per phase. Regarding general performance over the course of time, MANOVA (analysis of variance) showed a significant main effect of the phase on the average number of puzzles solved, $F(2,54) = 57.5$, $p < .001$, see Figure 3.9. The number of puzzles solved improved in later phases, clearly indicating a general learning effect. There was a main effect that justly did not reach significance of interface style on the overall number of solved trials, $F(1,27) = 3.4$, $p < .08$. Internalization subjects solved more trials (M=7.4, SD=.8) than the Externalization subjects (M=6.3, SD=2.1). No interaction effects were found.

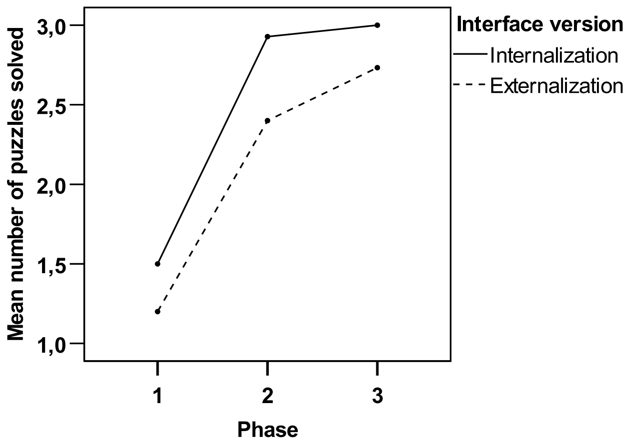



Figure 3.9 – Mean number of puzzles solved per phase and version

**Time needed for the puzzle-trials**

The time needed to solve the tasks, could logically only be calculated for trials that were actually solved. Only 22 of the 29 subjects had solved at least one task in each phase. MANOVA showed a main effect for phase on average solving time, $F(2,42) = 35.16$; $p < .001$. The time subjects needed to solve puzzles lessened in later phases. Post hoc comparisons showed that all subjects needed significantly more time in phase 1 than in phase 2, and also more time in phase 2 than in phase 3. The graph in Figure 3.10 suggests that Internalization subjects needed more time in the first phase. The effect of interface style was not significant overall, but a tendency could be seen, $F(1,21) = 2.79$, $p < .1$.

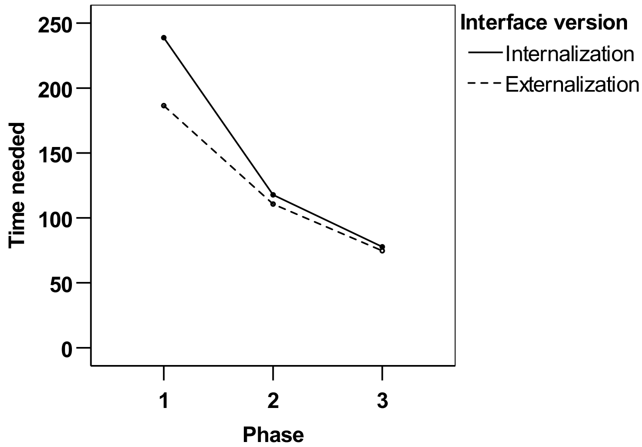



Figure 3.10 – Average time needed per solved trial, per version

Looking closer, it was only in the beginning, in phase 1, that Internalization subjects needed more time (M = 238.77, SD = 62.20) to solve the puzzles than Externalization subjects (M = 186.77, SD = 67.08). This difference was significant, $t(21) = 1.94$, $p < .05$ (one-sided). After this first phase, the differences were small and no longer significant. This is accordance with our hypotheses, stating better performance for the users of the Externalization interface, but only in the beginning.

**Superfluous moves**

Since we know with how many moves the puzzles could be minimally solved we can look at the deviations from this path, reflecting *extra, unnecessary* moves from the path leading to an eventual solution. In each of the three trials, the starting situation differed slightly (see Figure 3.3). The minimum number of moves to solve the first, second and third trial in each phase were 33, 31 and 30 respectively. Above, the results are reported concerning the number of solved trials and we saw that the number of solved trials increases in time. The downside of this phenomenon (not all tasks being solved in the beginning) is that certain analyses are only possible for tasks that were actually solved. Repeated measures analysis over the averages of the 3 phases requires that the subjects that can be included had to have at least one solved trial in every phase. This leaves us with substantially smaller groups to compare, as only 22 of the 29 subjects had solved at least one task in each phase. Repeated measures analysis with these 22 subjects showed no significant main effect or interaction effects for interface style on the number of superfluous moves needed for the trials that were solved. Nevertheless, we present the graph that belongs to this measure. When looking at it, attention is drawn at the relatively large difference in phase 3.

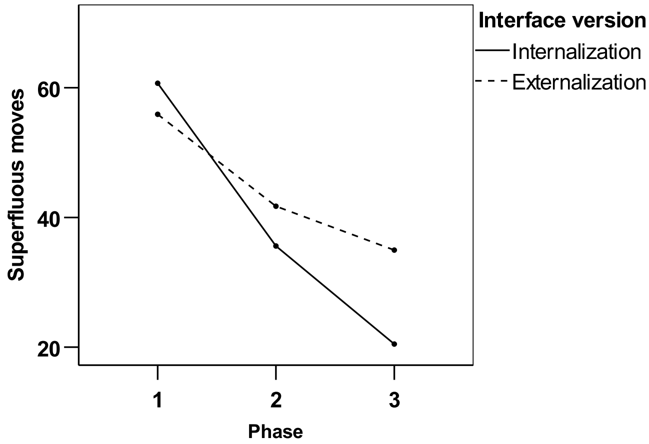



Figure 3.11 – Mean number of superfluous moves per phase and version

To be able to have more subjects in our analyses we decided to locally (per phase) compare the averages. It showed that in phase 1 and 2 there were no significant differences between the two interface styles, as the graph in Figure 3.11 also suggests. However, after the distraction task, in phase 3 there was. In this analysis already 28 of the 29 subjects could be included. At that point, the Internalization subjects significantly $t(27) = 2.04$, $p < .05$ outperformed ($M = 18.1$, $SD = 12.7$) the Externalization subjects ($M = 47.3$, $SD = 50.2$).

**Dead-end-states**

We introduced the measure “dead-end states” to inform us of how subjects behaved in terms of the insight they had, the “smartness” of their route. These are puzzle states that are off a direct solution path (see grey boxes, Figure 3.3), indicating trial and error search and lostness. Elaborate planning as is assumed in the Internalization condition is expected to cause subjects to reach fewer dead-end-states, and less frequently so. Repeated measures ANOVA showed an interesting nearly significant main effect of interface style (Figure 3.12). Overall, Externalization subjects reached more dead-end states ($F(1,28) = 3.58$; $p < .06$). In addition, there is a weak trend for an interaction effect of phase and version ($F(2,56) = 2.11$; $p = 0.13$). Internalization subjects significantly improved from phase 1 to 2 ($M = 11.4$, $SD = 4.70$ and $M = 3.4$, $SD = 3.18$), $t(14) = 5.80$, $p < .001$, one sided. They also improved from phase 2 to 3, nearly reaching floor level ($M = 3.4$, $SD= 3.18$ and $M = 1.47$, $SD = 2.07$), $t(14) = 1.96$, $p < .05$, one sided. Externalization subjects improved from phase 1 to 2 ($M = 12.67$, $SD = 6.91$ and $M = 6.47$, $SD = 7.11$), $t(14) = 2.74$, $p < .05$, one sided. But after the interruption it was different. From phase 2 to 3 they did not improve more. On the contrary, in phase 3 they reached more dead-end-states than before, although not significantly so. Here in phase 3 the difference between Internalization ($M = 1.47$, $SD = 0.53$) and Externalization ($M = 7.73$, $SD = 10.64$) was significant, $t(28) = -2.24$, $p < .05$. Externalization subjects reached more dead-end-states on average in phase 3 than Internalization subjects.

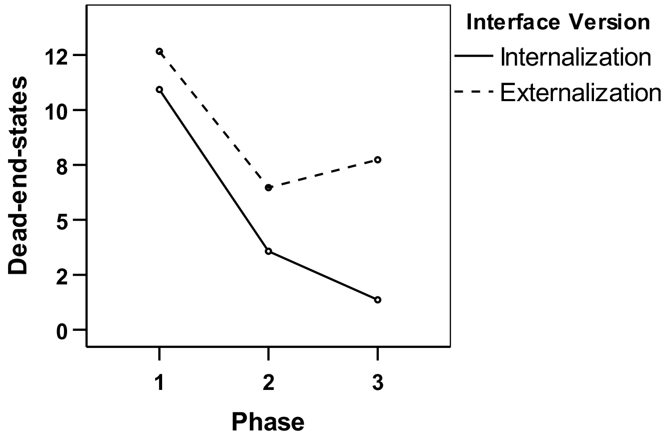



Figure 3.12 – Mean number of dead-end-states per phase, per version

**Number of moves performed in all trials, regardless of whether solved or not**
Above we made clear why it was not possible to always include all the subjects in the analysis. Still, what we *can* do is look at the number of actions users performed with the two interface styles, regardless of whether the problems were solved or not. We thus look at the influence of interface style, on the behavior in terms of number of actions performed. There was a tendency of interface style on the number of actions performed, $F(1,27) = 2.92$, $p <. 10$). On average, the Internalization subjects made fewer moves when trying to solve the puzzles, indicating that they spent more time and effort on working on a proper strategy. We take this as an indication that the Externalization version invited more trial and error behavior. Although the interaction between phase (1, 2, 3) and interface style was not significant, one can see in Figure 3.13 that after the distraction task between task 2 and 3, the Internalization subjects continued making fewer moves on average than before. With the Externalization subjects, this was not the case.

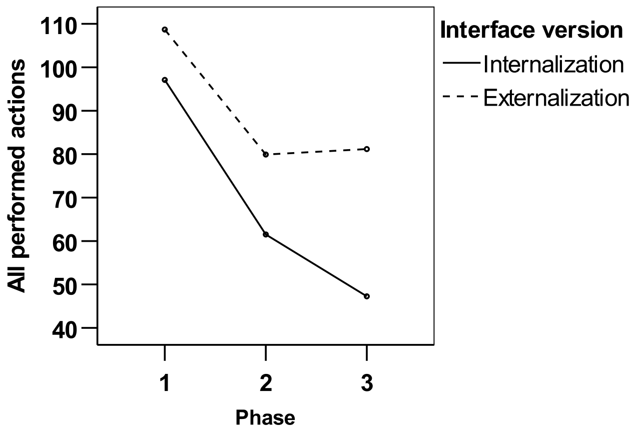



Figure 3.13 – Mean total number of performed moves per phase and interface style

**Consulting the rules**
Although subjects were able to consult the rules, this was not significantly different between the two versions. Whether they were consulted or not, and how often and how long, had no effect on any performance measure.

**Knowledge test and answers on the questionnaire**

The knowledge test that subjects received after the experiment consisted of several procedural knowledge questions and one explicit knowledge question. The Balls & Boxes puzzle contained more than 30 legal states (Figure 3.3). There were 7 procedural implicit knowledge questions about 7 of those states (open and multiple choice) in which subjects were visually presented with a puzzle state. They judged whether (or not) certain moves would lead to the solution, and had to explain why. The scores on these questions were high for both Internalization ($M = 6.3$, $SD = .96$) and Externalization subjects ($M = 5.8$, $SD = 1.20$), considering that the maximum score was 7. Even though it looks like a ceiling effect, there was still a trend that Internalization on average scored higher than Externalization on the procedural knowledge questions, $t(28) = 1.17$, $p < .12$).

To test explicit rule knowledge, the possibilities were limited. The puzzle had only few rules, and most of them were easy to grasp and remember. Rule 3 is the most crucial and defines the difficulty of the puzzle. Subjects were asked about this rule with a multiple-choice question. All Internalization subjects answered this question correctly, whereas of the Externalization subjects only 60% answered it correctly. This difference was significant ($Phi =-.50$, $p < .01$). This could indicate that often Externalization subjects reached the solution by trial and error (which is possible if one exercises some patience) without really knowing the rule. Furthermore, subjects had to estimate the clarity of the rules. They rated the question "the given rules were clear enough to solve the problem" (score range 1-5). Internalization subjects found the rules clearer than the subjects from the Externalization condition ($M = 4.13$, $SD = 0.52$ and $M = 3.53$, $SD = 1.25$), $t(28) = 1.72$, $p < .05$ (one-sided).

### 3.3.7 Discussion Experiment 1 session A

In general, we saw that performance over time of all the subjects increased in terms of the time needed for it, and the number of puzzles solved. Our first hypothesis stating that *initially* Externalization yields better performance was only marginally supported: there was the tendency that the Internalization took slightly *more time* in the first phase. The Externalization of information as we implemented it did *not* result in better performance in terms of *puzzles solved* in the beginning, as we expected. It was the other way around, we found the tendency that overall, the Internalization subjects performed better in terms of number of puzzles solved, regardless of the phase. This supports our general idea that guidance does not necessarily aid the user in the problem process, and that having to internalize rules and develop strategies by oneself yields better performance. CLT mentions Germane Cognitive Load as contributing to effort and motivation, and as such being beneficial for schema construction and learning. The finding that Internalization subjects (guidance was absent) took more time in the beginning, combined with the fact that they solved, fits nicely with this idea.

The second hypothesis stating that Internalization yields better performance in a later stage was supported in several ways. We saw that the time subjects needed stabilized. The tendency that Internalization subjects solved more trials persisted, regardless of the phase they were in. We also looked at performance in a more delicate manner, not in terms of time or number of puzzles solved, but at *how*, via which route, subjects reached their goal. We introduced the measure dead-end-states informing us about how subjects behaved, in terms of the insight they had, the "smartness" of their route. Stranding at dead-end-states is *not* what high Germane Cognitive Load should result in; it rather corresponds with

Extraneous Cognitive Load. Reaching them is an unwise and unnecessary thing to do, and if one would think ahead, one would reach those less often. We assumed that Internalization subjects do some smarter, more elaborate planning, while Externalization subjects are expected to solve more by trial and error and on the basis of interface cues. It showed that Internalization subjects performed better after the pause in which a distracting task was done and reached those dead-end-states almost significantly less often in all the three phases (indicating less lostness). Furthermore, there was also the trend-like interaction that after the interruption, Internalization subjects kept improving, while Externalization subjects fell back, reaching more dead-end-states than they did before. This confirms our expectation that after a delay during which they had another task, Internalization subjects continue to work on the basis of a more plan-based strategy as they did before. Externalization subjects, on the other hand, did not perform better after the interruption. They fell back depending on the interface, indicating a less elaborated plan. This might be caused by (too) high Extraneous Cognitive Load, which did not permit Germane Cognitive Load to be high, in turn leading to lower effort aimed at meaningful cognitive processes.

Because of the nature of the puzzles and the imposed timeout (the fact that the puzzle stopped after 7 minutes, for practical reasons), there were unsolved trials in all of the subjects. Our assumptions regarding Externalization involve also the fact that we expect more trial and error search together with less contemplation in general. With this in mind it made sense to look at how users use the interface and to focus on the amount of actions (clicks) they perform. In attempting to solve the problem, is smart to take time to analyze the problem and consciously work on a strategy. It is not smart to aimlessly click around and see if one reaches the solution by accident. We expected that the Externalization interface, because of its seemingly guiding features, was more inviting to users to behave in that way, causing too high extraneous (unnecessary) cognitive load. Again, a tendency was found here. Externalization subjects perform more actions than Internalization subjects do. This is in line with our assumption that more trial and error behavior takes place, and more actions, more clicking points at problem solving in a more display-based manner than problem solving on the basis of an internal plan. Also here, just as with the dead-end-states measure, after the pause the Internalization subjects kept on diminishing the number of actions, whereas the Externalization subjects just kept on behaving as they did before the pause.

The third hypothesis, in which we expected that Internalization would result in having better knowledge, was supported. We assumed that Internalization subjects, who could rely less on interface information, had to build a stronger, more elaborate plan (caused by high Germane Cognitive Load). When testing implicit knowledge both groups scored equally high, with a tendency that Internalization subjects scored higher. However, concerning the most crucial rule knowledge for these tasks (rule 3, stating that blue balls can never outnumber yellow balls) there was a significant difference. All the Internalization subjects could answer this question correctly, whereas only sixty percent of the Externalization subjects could - in spite of having the rules readily available for consultation. This serves our assumption that trial and error problem solving occurred more in the Externalization condition, provoked by the seemingly guiding nature of the interface. We interpret this finding as indicator of better understanding in the internalized condition. Lastly, the finding that Internalization subjects found the rules clearer than the

Externalization subjects did also fits well with the idea of better general understanding in the Internalization condition.

One of the reasons to adopt CLT as our theoretical framework was that we assumed that human behavior is not always as straightforward, consistent, and logical as it could be. This assumption seems justified, regarding the fact that in the 9 trials, sometimes solved trials were followed by unsolved trials (while the problem was the basically the same). The idea that the Externalization interface discourages devoting resources to beneficial processes (Germane Cognitive Load), is strengthened by the fact that Externalizations subjects were not very good at recalling the most stringent constraint of the problem (rule 3), indicating that more trial and error problem solving must have been taking place. Lastly, there was also the tendency that Internalization subjects rated the *clarity* of the rules higher. This is intriguing, because in the Externalization version of the puzzle subjects had interface feedback *and* were able to consult the rules. Internalization subjects, who *only* had the rules and no interface-help found the rules clearer. This could point at Internalization subjects making more effort to make sense of the rules and applying them, rather than just reading them and continue the problem in a display-based manner.

## 3.4 Experiment 1 session B: Performance after a considerable delay and transfer

We were curious to see how stable the results are from session A, which pointed at Internalization leading to more plan-based problem solving. The Internalization interface resulted in more solved trials, and it showed that these subjects roamed the problem space less in terms of actions performed and stranding at dead-end-problem states. In addition, the most crucial rule to the problem was better remembered by the Internalization subjects. Session B is a follow-up experiment conducted after a delay of eight months. We attempted to arrange the same subjects of experiment 1 session A, however only 14 of the 30 subjects were still available. To see whether the better knowledge as measured in the Internalization subjects had endured, we asked subjects to solve B&B again five times. Secondly, to see whether the better knowledge might result in better performance on a transfer task, we also confronted subjects with a transfer problem. Transfer problems require subjects to apply acquired skill on a different task of the same nature. To be able to measure differences in performance between the two initial groups (Internalization and Externalization) we presented all subjects with the same material this time.

### 3.4.1 Hypotheses

**H1: Internalization subjects will still have better memory of the rules and solutions**. They will be faster in recollecting knowledge needed to solve the puzzle and perform better.

**H2: Internalization subjects will perform better on the transfer task**

After five times B&B, we expect the two groups to perform more or less at the same level on *that* task. But when confronted with the transfer task (that has similarities, but also a few differences) we expect Internalization subjects to perform better, again because they possess better knowledge of the rules, and are still better able to apply it elsewhere.

### 3.4.2 Materials

**Balls & Boxes to test retention after a large delay**

With the visual rotation task of experiment 1, session A, attempted to mimic a small delay. Here we will also investigate whether the interface style one uses to learn and perform the task has an influence on remembrance of task rules and solutions, and consequently on performance after a *large* delay. To test this, after several months the same users will be presented with the same Balls & Boxes puzzle they worked with before. We will look at how well subjects remember task information and consequently perform after this delay. However, this time all subjects will receive the same interface style, the Externalization version. The idea behind it is that they all should have the same material to work with. We assume that the Internalization subjects from before have no problem with visual cues being presented in the interface, whereas this was absent earlier. The other way around, Externalization subjects might be at disadvantage when confronted with an interface that did not "help" them. To test knowledge retention for Balls & Boxes, all subjects solved the B&B puzzle in the externalized version five times. This time it will not possible to consult the rules anymore.

**Material to test transfer: The classic Missionaries & Cannibals puzzle**

We are interested in whether the interface version worked with before influences how well the acquired knowledge and practice are transferred to a similar transfer problem. To study this, subjects will be presented with another puzzle of the same M&C family, but with different characteristics. We will use a quite literal version of Missionaries & Cannibals (M&C), ironically the original version of the M&C as shown in Figure 3.1. This game, a Macromedia Flash Game (freeware by Plastelina Interactive ©) literally shows missionaries, cannibals, a river and a boat, all the semantic information that we tried so hard to get rid of in Balls & Boxes. The solution algorithm to this problem was the same as B&B, but there were some difficulties that did not exist before. The most important one is that there were not 5 of each entity as before, but 3, making it different at least at face value, and the playing direction was always from right to left. In addition, unlike in B&B, when users attempt illegal actions in this game, subjects would "die", as in a computer game, and the game would start over. Errors thus had more severe consequences, and subjects needed to exercise caution. Since errors are always being made, our subjects had to solve the puzzle as many times as they could in 8 minutes. So formally, this is transfer to both a different (but equivalent) task *and* a different interface at the same time.

### 3.4.3 Subjects and Procedure

Fourteen of the thirty initial subjects participated for the second time, and again they received a €5 reward. Fortunately, they came in the right numbers: 7 of these had worked with the Internalization interface, and 7 with the Externalization interface in session 1A.
The procedure was as follows:

1. Balls & Boxes (5 trials). The maximum time for each trial was set at 7 minutes. Slightly different starting-situations of the puzzle were used to avoid subjects simply repeating actions
2. Missionaries & Cannibals (solve as many trials as possible during 8 minutes)

### 3.4.4 Results Experiment 1 session B

**Balls & Boxes**

Encouraging results were found concerning solving the puzzle correctly again the first time. After not having seen the puzzle for eight months, it took the Internalization subjects only half the time that the Externalization subjects needed to solve the first B&B puzzle ($M = 432$ sec, $SD = 314$ and $M = 778$ sec, $SD = 397$). This difference was significant, $t(12) = -1.81$, $p < .05$. There were no further significant differences between the two groups concerning other measures. After the first puzzle, as expected, all subjects solved the remaining 4 trials of B&B puzzle equally well.

**Transfer task: Missionaries & Cannibals**

The graphical characteristics of the M&C puzzle differed considerably from B&B. Still the algorithm to solve it in itself was similar but the number of creatures to transport, and also the maximum number of creatures allowed in the boat was different. The same basic concept had to be applied to a situation that differed at face value. Of our 14 subjects, 10 managed to solve this puzzle one or more times. Just as in the B&B puzzle, Internalization subjects solved it the first time faster ($M = 176$ seconds, $SD = 72.5$ vs. $M = 302.8$, $SD = 202.7$), though it was just a trend, $t(8) = -1.44$, $p < 0.10$ (one sided). Moreover, on average Internalization subjects managed to solve the puzzle 3 times more often ($M = 4.83$, $SD = 1.94$) in the 8 minutes than Externalization subjects ($M = 2.5$, $SD = 1.73$) within the available time. This was significant, $t(8) = 1.94$, $p < .05$ (one sided).

### 3.4.5 Discussion Experiment 1 session B

Eight months is a long time, but the interface style subjects received in experiment 1 session A showed to make a difference. After session A, subjects were *not* told that they would be asked to solve the problem again after a long delay, and when we approached them for session B we explicitly told them it was not the same experiment as the one they had participated in before (so they were not encouraged to think about the experiment and the task again). Upon being confronted with the puzzle again, the subjects had to recollect the rules and solution strategy from their long-term memory.

The two groups both worked with an Externalization version this time and had exactly the same information from the interface. We looked at how long it would take subjects to remember and solve their first puzzle task correctly. It showed that Internalization subjects were indeed significantly faster (almost twice as fast). After that first success, the performance of both groups equaled over time. The above finding supports the hypothesis stating that Internalization subjects would still have a better memory of the rules and solutions of the puzzle.

The M&C puzzle that subjects received was also the same for everyone. Also here there were interesting results. The subjects that had worked with the Internalization version of B&B managed to solve this puzzle three times more often than Externalization subjects did, and this was significant. Furthermore, just like in B&B, Internalization subjects needed less time to solve the problem for the first time, although this was just a trend. It supports the second hypothesis stating that Internalization subjects will perform better on a transfer task.

In summary, we demonstrated that the interface style subjects worked with eight months ago still appeared to be of influence months later, both in solving the same puzzle again as in solving a transfer puzzle. We take it as encouraging support for the better memory of knowledge provoked by the interface style subjects worked with. Having to internalize information and not being distracted by externalized interface, allows users to be focused and concentrated on the primary task: studying the situation. One needs a strategy, which in B&B starts with realizing that one has to start with the "bad" entities (here the blue balls), to solve the puzzle. Connecting to CLT, having to internalize certain information oneself and not being bothered by interface information keeps the problem solver proactive. This imposes, we assume, only low Extraneous Cognitive Load and stimulates and allows high Germane Cognitive Load in the user. The latter has a positive effect on not only performance initially, but also leading to better remembrance over a long time.

## 3.5 Experiment 2: More Balls & Boxes, interface style and the influence of instruction

In the preceding sections, we elaborated upon the first experiments with the Balls & Boxes applications. Our hypotheses concerning certain measures could be accepted, even though some of the results were solely backed up by statistical tendencies, rather than by significant results at a conventional level. However, the patterns in which behavior occurred strengthen us in our assumption that the Internalization condition by itself provokes more metacognitive activity, and especially planning.

So far, we focused on, and theorized about mental effort and plan construction provoked by the interface itself. In the light of CLT, it was Extraneous Cognitive Load that we tried to keep low, thus permitting Germane Cognitive Load (associated with mindful effort) to be high. However, in the current experiment we will attempt to provoke mental effort also from the outside. *Externally* induced effort, a level of effort instigated by external instruction could very well have an influence on performance. What will be subjects' behavior be like if they are not just confronted with one of the two interface styles, but also explicitly instructed to plan moves carefully, versus shallowly solving the tasks? To see whether external instruction before starting the tasks has an influence, and whether this has different effects on users of the Internalization and Externalization interface respectively we will devise another experiment with the same material. It is the interaction between interaction style and planning instruction that we focus on. It might be the case that the negative effects that we saw of the Externalization version are lessened by explicitly stimulating subjects to think deeply. However, when Externalization subjects are instructed to approach the problem in a shallow manner, they might perform even worse than without instruction. Within the Internalization subjects, whose intrinsic motivation is thought to be higher already because of the thought provoking qualities of the interface, instruction might have an influence, but perhaps to a lesser degree.

### 3.5.1 Material

As in experiments 1 session A, we used the Balls & Boxes application. Because we had seen in experiment 1 session A that our subjects (in the first session) failed to solve quite some trials, we decided to let the puzzles start easier and change over time. We constructed

3 puzzle versions in which the number of balls in the puzzle gradually increased. The first puzzle was easier in terms of complexity (Figure 3.14). Furthermore, the timeout for the puzzles was increased to 10 minutes. The solution algorithm is basically the same, but more balls make it *look* more difficult, and the number of ball-moves increased considerably, the number of dish-moves not so much (see end of section).

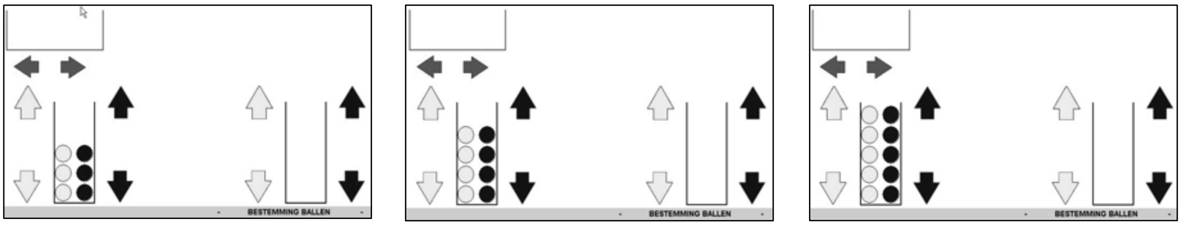



Figure 3.14 - Increasing number of balls in B&B3, B&B4 and B&B5
(Internalization version)

Only the constraint of the number of balls that could be transported had to be lowered in B&B3, if not the problem would be extremely easy to solve (all the bad ones could be transported at once, the problem would not be the problem anymore). What was also different with experiment 1 is that the rules of the puzzle were not consultable anymore. The idea behind it is that in this manner, making inferences about these rules is encouraged (in both conditions).
**B&B3**: 3 blue / 3 yellow balls. 2 balls fit in dish. Solution 29 moves (11 dish-moves / 18 ball-moves)
**B&B4**: 4 blue / 4 yellow balls. 3 balls fit in dish. Solution 27 moves (9 dish-moves /18 ball moves)
**B&B5**: 5 blue / 5 yellow balls. 3 balls fit in dish. Solution 33 moves (11 dish-moves/ 22 ball moves)

### 3.5.2 Subjects and Design

Thirty-one subjects, aged 19-35 were randomly assigned to four conditions (Table 3.1). They were undergraduate students from Utrecht University and received a €5 reward afterwards. Our 2x2x3 design has two between-subject independent variables: interface style (Internalization or Externalization) and planning instruction (low or high, see section 3.8.4). The independent within-subject variable was "Phase"; there were three phases with in each a different puzzle version (B&B3, B&B4 and B&B5).

Table 3.1 – Design experiment 2 and number of subjects per condition

| | Low planning instruction | | | High planning instruction | | |
|---|---|---|---|---|---|---|
| | Phase 1 | Phase2 | Phase3 | Phase 1 | Phase2 | Phase3 |
| Internalization | | 7 | | | 8 | |
| Externalization | | 8 | | | 8 | |

### 3.5.3 Measures

The dependent performance measures variables were (all logged by the computer):

- the number of solved puzzles

- time needed to solve the puzzles
- superfluous moves: the deviations from the shortest path
- the number of illegal moves
- dead-end-states (only possible in B&B5, in phase 3)
- questionnaire: in the knowledge test after the trials we measured how well subjects had learned the rules and Likert-scale questions concerning perceived amount of planning (see Appendix A)

### 3.5.4 Procedure and Instruction

Subjects solved 7 trials (3xB&B3, 2xB&B4, 2xB&B5, it took about half an hour). After completing them, they were presented with the knowledge questionnaire (10-15 minutes). At the start of the experiment, subjects received a general instruction on the course of the experiment, starting with a screenshot and the phrase "*All balls should be transported from one side to the other. However, there are constraints, not everything is allowed. Find out for yourself how it works*." After this, our independent variable "planning instruction" was applied:

Low planning instruction:
*"Try to solve the puzzle as fast as possible, making mistake is not a problem. Good luck!"*
High planning instruction:
*"Try to solve the puzzle as economically as possible. Think hard, plan with care, it pays off. Good luck!"*

### 3.5.5 Results Experiment 2

**Solution Times and number of puzzles solved**

The time that subjects needed to complete the puzzles was influenced neither by interface style, nor by planning instruction. This time, in contrast to experiment 1 session A where we *did* find a tendency for the number of puzzles solved, we found no significant effects here ($p > .05$) Almost all the subjects managed to solve all the puzzles, for which the reasons can be twofold. Firstly, B&B3 looks and seemed to be easier, although solution-wise it is not so different from B&B5 that was used earlier. It might be the case that starting with B&B3 and then increase the number of balls is actually a good training sequence. Secondly, the timeout was set at 10 minutes instead of six in experiment 1, so subjects had more time. The fact that almost all the trials were solved by all subjects does have the advantage that a repeated measure analysis with almost all subjects is possible, in contrast with experiment 1 session A where we could not do this.

**Superfluous moves**

Just as in experiment 1 we knew in how many moves the puzzles could be minimally solved so we can look at the deviations from this path, reflecting extra, unnecessary moves from the path leading to an eventual solution. The numbers of moves, the solution lengths were 29, 27 and 33 moves for B&B3, B&B4 and B&B5 respectively (dish-moves and ball-moves together). In experiment 1 session A the problem was that subjects could only be included when they had solved at least one trial in every phase, leaving us with substantially smaller groups to compare, as only 22 of the 29 subjects had solved at least one task in each phase.

Here however, most of the tasks were completed because of the earlier mentioned reasons, and therefore we could include all 31 subjects in the analysis. Because of the fact that the number of balls that had to be transported differed between the different phases, the scores have been transformed to z-scores (Figure 3.15 and Figure 3.16). Repeated measures analysis showed an overall main effect (almost significant at a conventional level) of interface style on the number of superfluous moves needed for the trials that were solved, $F(1,27) = 3.22$, $p < .08$. The Internalization subjects outperformed the Externalization subjects by making less superfluous moves on average over the three phases ($M = -.32$, $SD = .76$ vs. $M = .13$, $SD = .58$, in z-values).

**Low-Planning instruction**

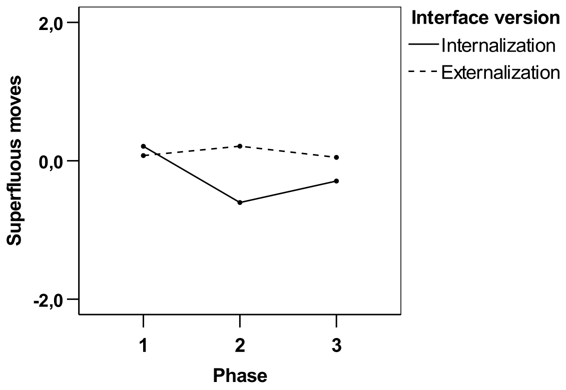



Figure 3.15 - Mean Z-values Superfluous moves in **Low-Planning** instruction per phase / version

**High-Planning instruction**

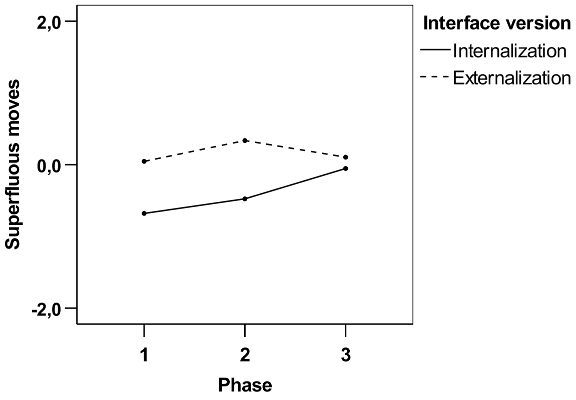



Figure 3.16 - Mean Z-values Superfluous moves in **High-Planning** instruction per phase / version

If one compares the two figures above, the patterns for the Externalization group are practically identical in Figures 3.15 and 3.16 (dashed lines). Externalization subjects were not influenced by planning instruction. A repeated measures ANOVA *within* the Internalization group (continued lines in the Figures 3.15 and 3.16) was performed. There was a significant interaction effect, $F(2,26) = 3.25$, $p < .05$ of puzzle phase and planning instruction. It showed that Internalization subjects were influenced by the planning instruction: in phase 1, subjects that received a high-planning instruction made considerably less superfluous moves. After phase 1, there were no differences anymore.

**Dead-end-states**

In this experiment, dead-end-states (see gray states in Figure 3.3) only existed in the problem space of B&B5. B&B4 and B&B3 did not have them, since they were simplified versions. B&B5 was presented in phase 3, but not after a memory-erasing task as was the case in experiment 1 session A, when difference in interface style started to have effects. In phase 3 with B&B5, there were no significant effects, nor interaction effects of interface style or planning instruction on dead-end-states in phase 3.

**Attempted illegal moves**

Results so far point out that Externalization subjects are not influenced by planning instruction, but Internalization subjects are. Some variables were only measurable in the internalized interface, e.g. “attempted illegal moves”. In the externalized version one could *only* make legal moves, while in the internalized version it was also allowed to *attempt* illegal ones, since the control arrows were always clickable. In experiment 1 session A it made no sense to look at this since there was nothing to compare, but now we have another between subjects variable, namely instruction. We analyzed the influence of instruction within the Internalization group on the amount of attempted mistakes. Within the Internalization condition we compared the number of times that this occurred, in relation to planning instruction. Because of the fact that the amount of balls that had to be transported differed to a great degree between the different phases, the scores have been transformed to z-scores.

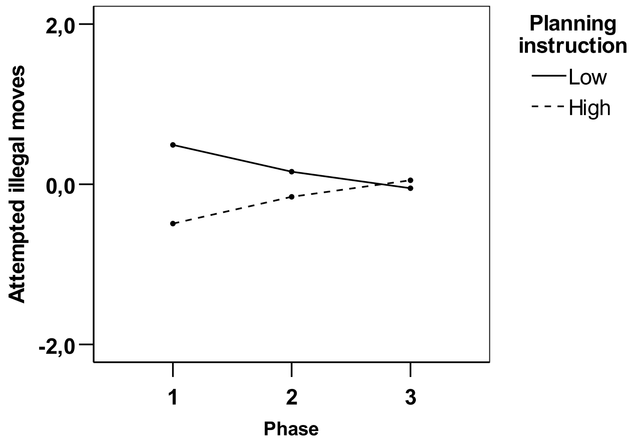



Figure 3.17- Mean Z-values number of attempted illegal moves by Internalization subjects per puzzle version

A repeated measures ANOVA showed an interaction effect $F(2,26) = 4.27$, $p < .05$ (Figure 3.17). In the Internalization condition, in the first puzzle (B&B3) the high planning instruction ($M = 81.50$, $SD = 43.94$) resulted in fewer attempted illegal moves than low planning instruction ($M = 163.75$, $SD = 96.05$), $t(14) = 2.20$, $p < .05$. If the instruction was to plan carefully, indeed subjects attempted substantially fewer illegal moves. In the Internalization condition, instruction did have its influence on behavior in the first phase. After the first phase, the difference gradually equaled.

**Questionnaire: Knowledge and Perceived amount of planning**

This time there were three questions concerning the crucial rules of the puzzle, the answer could be correct or incorrect (total score range 0-3). An ANOVA showed, as in earlier experimentation that the knowledge of the rules afterwards was influenced by interface style, $F(1,27) = 10.03$, $p <. 05$. Planning instruction had no influence. As before, the average knowledge acquired by Internalization subjects ($M = 1.95$, $SD = .77$) was better than for the Externalization subjects ($M = 1.0$, $SD = .82$). Since planning was what we tried to influence with our instructions, we also asked subjects to rate the amount of planning as perceived by themselves on a 5-point Likert scale. An ANOVA yielded just a tendency of an interaction of planning with interface style which did not reach significance, $F(3,27) =$

2.24, $p < .14$. However, when looking at the pattern, it fits earlier mentioned results quite well; only in the high planning instruction, interface style has an effect. Within this high planning condition, subjects that worked with Internalization ($M = 3.38$, $SD = 0.92$) felt that they planned more than subjects in the Externalization condition ($M = 2.38$, $SD = 1.06$), $t(14) = 2.02$, $p < .05$.

### 3.5.6 Discussion Experiment 2

We analyzed the influences that interface style (Externalization or Internalization) has on performance and knowledge acquisition in an earlier used problem solving task. In this experiment we varied the planning instruction (low or high) that subjects received to see whether this instruction has an influence on performance, but especially whether the planning instruction has different effects in the two interface styles. Subjects were either encouraged to do the task fast, errors did not matter (low planning) or to plan carefully and work as efficiently as possible (high planning). The puzzles were different in the sense that they started easier, and gradually increased in difficulty. We assume that this fact caused that almost all the tasks were solved by all the subjects, regardless of interface version or instruction.

The planning instruction subjects received *alone* had no influence on behavior and performance. Also instruction had no influence on the performance measures *number of puzzles solved* or *solving time*. In addition, the interface version by itself had no influence on time needed (as in experiment 1 session A). However, unlike in experiment 1 session A, also the number of solved trials did not differ, due to the fact that they started easier and the larger timeout of the software.

Still, we can study how subjects reach their solutions, in terms of directness or efficiency. Superfluous moves, the extra unnecessary moves aside an ideal path is a measure we used as an indicator of planning by subjects (for a similar measure, see O'Hara & Payne, 1998). This measure focuses on *how* subjects solve the problem, not on *if* or *how fast*. It reflects the directness, the efficiency of the path that subjects chose. Less, or even better, no "wandering back and forth" at all in the problem space, is taken as indicator of planning and contemplation by subjects (corresponding with high Germane Cognitive Load). The opposite would be just trying to solve the problem by trial and error, making many unnecessary extra moves, as can be expected when extraneous CL is high. Concerning extra moves, Internalization subjects were positively influenced by the planning instruction and Externalization subjects were not. This effect was the largest in the first puzzle. Here, high-planning Internalization subjects outperformed the three other groups by needing far fewer extra moves, thus displaying smarter, more thoughtful behavior. Subjects acted more carefully and considered their moves more. It seems that being confronted with the Externalization interface makes a subject ignore, or even forget the planning instruction all together. The findings are confirmed by subjects' own judgments of their planning; only subjects in the internalized interface with a high planning instruction reported having done a considerable amount of planning. The fact that subjects' own judgment of planning coincided with superfluous moves findings indicates that superfluous moves are an adequate measure for planning.

Only the interface where subjects had to internalize the needed information allowed attempts of illegal moves (besides the mentioned *legal* superfluous moves), since all controls were always clickable. The interface gave little information away about the rules of the puzzle, so it was no surprise that subjects in the Internalization condition attempted illegal moves at some point. This happened in the Externalization condition also, but our application was not able to log these events. Subjects with low planning instruction attempted twice as many illegal moves as the ones with high planning instruction. This again reconfirms that in the Internalization condition, planning instruction has an influence. High-planning instruction facilitated smarter behavior, and low-planning instruction inhibited planning resulting in visiting more dead-end-states.

After finishing the tasks, the knowledge of subjects was tested, and proved to be influenced *only* by the interface style, as we saw before. Although planning instruction in one condition resulted in fewer extra moves and higher perceived planning, it had no influence on how well the knowledge was remembered. One might expect that the high planning instruction *and* having to internalize information would re-enforce each other, resulting in even better knowledge, but this proved not to be the case. Interface style alone still was the main convincing factor of influence. Perhaps the nature of the puzzle was such that all Internalization subjects acquired the knowledge as good as it can be already, and that planning instruction (and consequently behavior) therefore could not make a difference anymore.

We have to conclude that in the externalized condition, subjects were “deaf” to the planning instruction. It was the combination of interface style and planning instruction that was deciding in subjects’ behavior. Externalization seems to encourage trial and error problem solving, which stands opposite to planning. Relying on interface information in this manner, making little use of learned knowledge corresponds with display-based problem solving behavior as defined by O’Hara and Payne (1998, 1999). In the Internalization condition, the given instruction was actually obeyed.

Regarding CLT in its original context (instructional science) this issue of “being deaf to instruction” is interesting. If during computer-based education in a given situation it is crucial or important that instructions are being followed, one might consider not using too much externalized information that is in the way of Germane Cognitive Load. It would be interesting to investigate what happens if the planning instruction is being repeated, to look at different levels of Externalization and feedback, and perhaps adaptively derive how well subjects are doing, and re-provide subjects with planning instruction based on that.

## 3.6 General discussion and conclusion Experiment 1 and 2

Our research questions concerned the influence that externalizing certain information on the interface, versus *not* externalizing (users have to internalize themselves) had on users’ plan-based behavior and consequently performance on problem solving tasks requiring planning. We investigated the conditions under which *externalizing* interface information by destination feedback influences strategy choice, and consequently performance. Furthermore, we investigated which interface style causes the best knowledge about the tasks, rules and solutions afterwards, the long-term effects after eight months, which interface style leads to the best performance in a transfer situation, and lastly at whether the influence of deliberately instructing subjects plan has different effects in the two interface styles.

In general, it was surprising for us to see that in our experiments so far, Externalization showed even fewer advantages than we thought, none to be precise. We expected Externalization at least to be of help in the beginning when users were not familiar with the system and the problem. It was, but only in time taken on the first few trials – and this was just a trend, that was not confirmed in the number of puzzles solved correctly. This very small advantage did not concur with the findings of Zhang and Norman (1994) and Zhang (1997). When our subjects worked with an interface where certain information had to be internalized by themselves, they were at advantage on several levels. *Not* having externalized information led to enduring advantages, whereas being guided by externalized information in the interface indeed had unbeneficial effects. The Internalization interface might not give the user the idea that he is being assisted, but causes more metacognitive activity such as planning and construction / application of smart strategies, corresponding with high Germane Cognitive Load as CLT mentions it. The result, consequently, is better performance (efficiency), and better imprinted knowledge about the task and its solutions.

In experiment 1 session A we saw the trend that Internalization subjects solved more trials, made less superfluous moves, reached fewer dead-end-states and performed fewer actions in the problem space in general. Also the most crucial task knowledge was remembered better by Internalization subjects. During experiment 1 session B that took place several months after, these same Internalization subjects still were at advantage. They were sooner able (almost twice as fast) to recollect the needed information to solve the same puzzle again for the first time, even though it involved the Externalization interface. They also managed to solve a transfer task faster for the first time, also indicating at better and more readily available knowledge of the task and its solutions. It is remarkable that the influence of working with one interface or the other has effects even after such a long time. Finally, they also managed to solve this task more often in the time given. Experiment 2 showed that the subjects working with the Externalization interface were deaf to instruction; instruction had no effect at all, how we instructed them before, made no difference. It was the Internalization condition that actually allowed the instruction to be followed. The Externalization interface suggests assistance and seems to draw all the attention away from the instruction and the task, which is exactly what corresponds with the notion of too high Extraneous Cognitive Load. The attention of users of the Externalization interfaces group is immediately taken by the guiding nature of the interface, and the feeling that the system is doing something for them makes these users less proactive.

Making more superfluous moves at various instances as happened in the Externalization condition indicating trial and error, is a behavior that fits with Extraneous Cognitive Load: actions that are unnecessary, and *not* contributing to mindful processing and development of strategies. Giving users the idea that they are being assisted or even that the task is being carried out partly for them, combined with the distracting features that Externalization of information can have, leads users to simply keep on solving without feeling a need to figure the problem out more deeply. Subsequently, less active learning takes place. This idea of attention taken by an interface fits with Carroll and Rosson's (1987) paradox of the active user – users of computer systems are so consumed with immediate productivity that they are not motivated to take time to learn better ways of accomplishing a task.

In the context of a cost trade-off, contrary to the Externalization group, the Internalization group was confronted with making "errors" (attempted errors) for which an explanation was not immediately available to them. Internalization subjects are not informed what is allowed with the controls, and sometimes attempted illegal moves. When they do so, they might realize that it is worthwhile to think a bit about why certain actions are not legal at the moment, realizing this pays off in the end. Externalization subjects were not confronted with actual mistakes (one could not make illegal moves, only inefficient legal moves were allowed), they simply kept on solving without applying meta-cognition (Tabachneck-Schijf, 1992). Attempting errors incurred a cost: a dialog box popped up which they had to click away, and their attempted move was reversed. Though both groups consulted the rules equally often, this cost probably contributed to motivating the Internalization subjects to study the rules better in order to avoid incurring the cost. We found that the Internalization group possessed better explicit knowledge of the rules and engaged in more planful problem solving. Applying more metacognition to avoid a cost concurs with the findings of O'Hara and Payne (1998, 1999).

We find the results so far encouraging in that lessening the amount of externalized knowledge apparently can encourage cognitive and metacognitive behavior. Therefore, we feel that the issue of manipulating the amount of Externalization in interfaces deserves more attention. Puzzles like B&B, TOH and the 8-puzzle that were used in mentioned research by others (see chapter 2), have been extensively studied in the problem solving literature, and user behavior can be precisely traced with them. They can be solved both by planning and by trial and error. Although being excellent tasks from a scientific point of view, a drawback is that findings from these experiments may not be relevant in richer, life-like tasks. O'Hara and Payne (1998) tried to examine whether their findings with an abstract problem (the 8-puzzle) could be generalized to an administrative task. In their experiment, subjects had to prepare letters that had to be sent to visitors of a conference, by copying the information and pasting it into the appropriate letter. Results showed that insights gained from experiments using an abstract problem, were also applicable to a less abstract administrative task. We argue that in problem solving situations where people need to learn the underlying rules of a system, make as little mistakes as possible, or find a solution as economically as possible, the use of a plan-based approach is valuable. If it is preferable that people use a plan-based approach during interaction, one should find out how people can be persuaded to use that approach. We will investigate our assumptions in the next chapter with a more realistic life-like task of which the functioning and appearance are more towards everyday-realism, and make a shift from classic puzzles to a task involving constraint-satisfaction scheduling. The difference is that not so much the "trick" to a math-like puzzle has to be found or that continuous action sequences have to be performed. We will extend our previous work in the sense that users are presented with tasks that have a "real life" goal, appearance, and controls. Not so much the specific functions or controls of the application have to be "learned" (they are quite straightforward) but users rather have to uncover the problem solving heuristics (provoked by different interface styles).

The findings so far allow us to address research questions 1, 2, 4 and 6 of our more specific research questions of section 1.5.1. Question 1 concerned how problem solving performance evolves over time. In both interface versions, subjects improved rapidly over time. There was the tendency that the Internalization subjects solved more puzzles out of the total. After a small pause, it showed that Externalization subjects reached more dead-end-states than before, and also more superfluous moves, both pointing at worse imprinting of knowledge and more trial and error behavior. Question 2 concerned transfer of skill to another task. It showed that Internalization subjects were better at this; they were better able to apply principles form the initial task to a transfer task. Question 4 concerned the influence of planning instruction on performance. Only Internalization subjects actually followed the instructions, and when induced to plan carefully, actually performed much better than the others did. In question 6, we wondered how performance would be after a large delay. Results showed that Internalization subjects were better at correctly recalling procedures and solutions after the delay.

In the next chapter, we elaborate on more experiments addressing the same, and also some other of the more specific research questions using a different task, an application that is more like a standard office task as in real work situations.

# 4 Evidence from more realistic material: Conference planner

In the preceding chapter, we studied abstract puzzle tasks, and found several pointers confirming our expectations (Van Nimwegen, Van Oostendorp & Tabachneck-Schijf, 2004c, 2004d). Nevertheless, despite the results, the tasks were abstract puzzles of a repetitive nature, and once one knew the solution, solving the next problem was always a similar procedure. Our puzzle task was not very much like computer tasks in daily life, and the question is whether our results can be generalized to tasks that are more realistic as well. The advantage of the Balls & Boxes task was its tractability and the amount of control one can exercise over it, the planning characteristics it has, and the fact that its original version (Missionaries & Cannibals) has been successfully used in other research (Ernst & Newell, 1969; Jeffries, Polson, Razran & Atwood, 1977). In real life, there are mostly more variables involved in tasks that are part of a real job, such as in for example spreadsheet, drawing, or other office-like applications where actions are less repetitive and more complex. Still, if one wants to conduct grounded empirical research, it is not feasible to *start* with scrutinizing on an application that is extremely complex, such as for example the complete Microsoft Office © package, or an entire operating system. Therefore, our next step up will be to study Internalization and Externalization in a more realistic task that is still tractable in terms of the problem space users have to navigate through (see section 4.2).

With experiment 3, we address questions 1 and 5 of the specific research questions (section 1.5.1). The first question (as in experiment 2) concerns repeated problem solving performance between the two interface styles on several measures. The fifth question concerns the attitude individual users have to problem solving in general. In experiment 2, we investigated the influence of *externally* induced motivation. Here, we focus on more stable personal qualities of people that could influence behavior. We wonder whether differences in cognitive style of users (*internally* determined task conception or problem solving attitude already present in people) influence behavior instigated by the two interface styles. In CLT, motivation is seen as related to Germane Cognitive Load (Paas, Renkl & Sweller, 2003). We theorized that Extraneous Cognitive should be low, to permit high Germane Cognitive load. However, if the latter is already high in users, perhaps this alone will already lead to more elaborate planning and better performance than users that have a low motivation for problem solving. It could be that the specific interface style has less, or no influence on subjects that have high intrinsic motivation for problem solving.

In experiment 4, using the same realistic application, we address questions 2, 3 and 7 of the specific research questions (section 1.5.1). Question 2 concerns transfer of skill to another task; we investigate whether the interface style one worked with before influences performance on another task (as in experiment 1B). With question 3 we investigate whether the interface style one worked with before influences performance when working with the other interface style afterwards. Question 7 was whether severe *interruption* in the problem solving process has different effects in different interface styles. Our assumption is that if actions are more planned, and if indeed knowledge and procedures are better represented in the Internalization condition, this will result in smoother and more efficient task resumption for users that had to internalize all the information by themselves.

## 4.1 More realistic material: A conference planning task

Among the properties of our new more realistic task, the most important one is that it requires planning. Furthermore, the planning should be of such a nature that when it would be applied properly, it would result in a more efficient solution. The task should not be extremely difficult, but such that applying a certain plan-based approach is necessary to solve the situation in the most efficient manner. We chose for a task that is aimed at exactly that: a constraint-satisfaction scheduling task that concerns planning *itself*. There are ample examples of software made for that, such as at university-class-planning or airport-gate-assignment-scheduling etcetera. The task we chose involves planning speakers that give lectures at a 1-day conference. The problem solving situation is as follows (see Figure 4.1). There is a conference to be held in a conference facility, and there is a list with a number of speakers that will give a talk that day. The conference facility has several auditoriums with differing features. Speakers have to be scheduled into a time grid over that day. Not all the timeslots in the grid are always available. Some of them where unavailable all the time, indicated with light gray, for example the timeslots during lunchtime (13:00), but also some arbitrary other slots (e.g. 10:00, room "Maxima"). We designed the problems such that a correct solution where all to be scheduled speaker fit in the grid always existed. The empty available timeslots were shown in white, and the ones that were already occupied by a speaker would display the name of a speaker. In Figure 4.1 the following is visible:

On the **right** side of the application, the time grid is displayed, with three conference rooms (in Dutch: Zaal): Beatrix, Maxima, Irene that each vary on the following variables:

| | |
|---|---|
| **Capacity:** | Different rooms fit different audiences<br>Zitplaatsen: 80, 101, 250 |
| **Equipment:** | Some rooms have projectors available, others do not<br>Beamer (Dutch/German pseudo-Anglicism meaning projector"): "Ja" (yes) or "Nee" (no) |
| **Timeslot duration:** | There are timeslots of 1 hour and of 2 hours<br>For example, the 09:00 slot in room Maxima is 1 hour; the 11:00 slot is two hours |

On the **left** side of the application the speakers (in Dutch: Sprekers) are listed with their names (in Dutch: Naam) and different varying demands regarding

| | | |
|---|---|---|
| **Capacity:** | Size of the audience | (in Dutch: Toehoorders) |
| **Equipment:** | Projector needed or not | (in Dutch: Beamer, Yes=Ja, No=Nee) |
| **Duration of talk:** | 1 or 2 hours | (in Dutch: Uren) |

When facing this task, one can take multiple (correct) approaches to schedule the speakers in such a way that they all fit. In the above example, speakers with an audience of 101 fit in a room of 250, speakers not requiring a projector (beamer) still can use a room with a projector and speakers that need 1 hour can be placed in a slot of two hours. It is important to realize that none of the latter situations are true the other way around. A large audience does not fit in too small a room, if a speaker needs a projector he cannot speak in a room without one, and a 2-hour lecture cannot be held in a 1-hour slot. If one tackles the tasks in a smart way, one should focus first on speakers with stringent constraints, e.g. in the above example:

- speakers that have an audience of 101 or more (they fit in only 1 room; they have to be placed there)
- speakers that do need a projector (2 rooms possible; they cannot be placed in room “Maxima”)
- speakers that need a slot of 2 hours (9 two-hour slots available; they have to have a 2-hour slot)

Unlike in tasks such as the Tower of Hanoi or Balls & Boxes, there are many more ways to optimally solve the scheduling task. A fixed order is not mandatory, but in practice, placing the speakers in a certain order will increase the chance of the most efficient solution. Even with more correct solutions existing, without some degree of planning, the scheduling will not be optimal and extra moves (corrections on the assignments so far) will be needed. If the entire situation is not taken into account, the participant will be stuck in some later phase of the task because (s)he will encounter a speaker that does not fit in any of the slots that are left. Only by planning, one will be able to fill the schedule without extra moves.

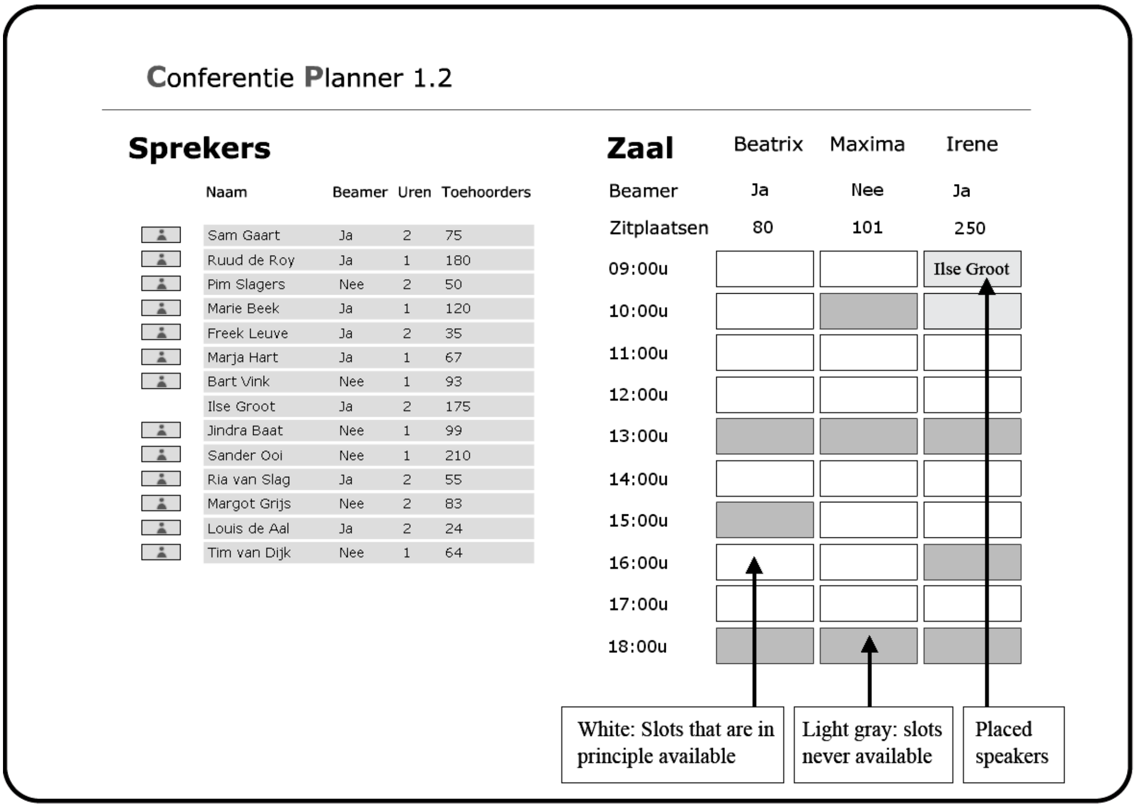



Figure 4.1 - The conference planner application

A useful heuristic would be to look for the speakers with the most constraints and place them on the schedule first, matching their constraints with the constraints of the schedule slots. It can also be effective to start with identification of the slots that are hardest to fill (the ones unavailable to speakers needing 2 hours, limited number of people, and no beamer) and then assign speakers to those slots. An “easy” approach that does not involve much planning would be to start from the top of the list and work your way down, filling up the different slots. However, this approach is not so smart: one will find that some speakers cannot be placed anymore, after which speakers must be shifted around to make room, requiring extra moves. One can imagine still other strategies that do not involve a lot of planning, but strategies that are more like a trial and error approach (such as random selection of speakers).

## 4.2 Material: The Conference Planner Application

We developed an Open Source software application called "Conference Planner" which simulated the planning of speakers for a conference. The software (implemented with Macromedia Flash MX ©) logged all the moves participants made. The "Conference Planner" was developed by The Open University of The Netherlands and funded by the European UNFOLD Project. It consists of four different components. The first one is the dynamic interface that shows each set of demands for a conference and allows the end-user to solve the problem in an easy way, based on drag & drop movements. The second one is the core of the application itself: the set of rules and related algorithms. Here is defined which actions are permitted, based on the requirements of the experiment and which are the subsequent consequences. The third component is a database, with all the scenarios used in the experiment. The fourth component is the logging-module that writes all the clicks and drag & drop moves and their associated times, as well as waiting times to an external spreadsheet. The logs provide data for analysis of the results.

### 4.2.1 Internalization and Externalization

Also of this task, two versions were constructed, one with an Externalization interface and one with an Internalization interface. The difference between Externalization and Internalization was implemented by highlighting legal slots in the Externalization condition (Figure 4.2) where a speaker can be placed. In the Externalization interface version when one clicks on the boxed icon on the left side of a speaker (in the list on the left) the legal slots (those satisfying the constraints and being available) in the timetable turn green. Note that this does not show the best or smartest slot to place a speaker, but simply which slots are possible. To move a speaker from the left to a slot on the right, the little boxed icon in front of each speaker's name had to be "picked up" and "dragged" to its destination slot with the mouse. Note that the behavior of the Externalization interface is slightly different compared to the interface of Balls & Boxes. In Balls & Boxes, the legal moves are visible all the time; one did not have to click to obtain this information. Here, the screen is as in Figure 4.2 and only *on click*, and *during dragging* a speaker the externalized information is visible. In the below example, when clicking on the icon left of "Ilse Groot" who needs 2 hours, a projector (beamer) and has an audience of 175, this speaker can be placed at only 4 positions that are all in room "Irene". In the *Internalization* condition, the destination feedback in green was absent. On clicking and dragging a speaker, no interface changes would occur, and one has to look up information and constraints by oneself all the time. No other differences existed between the two conditions.

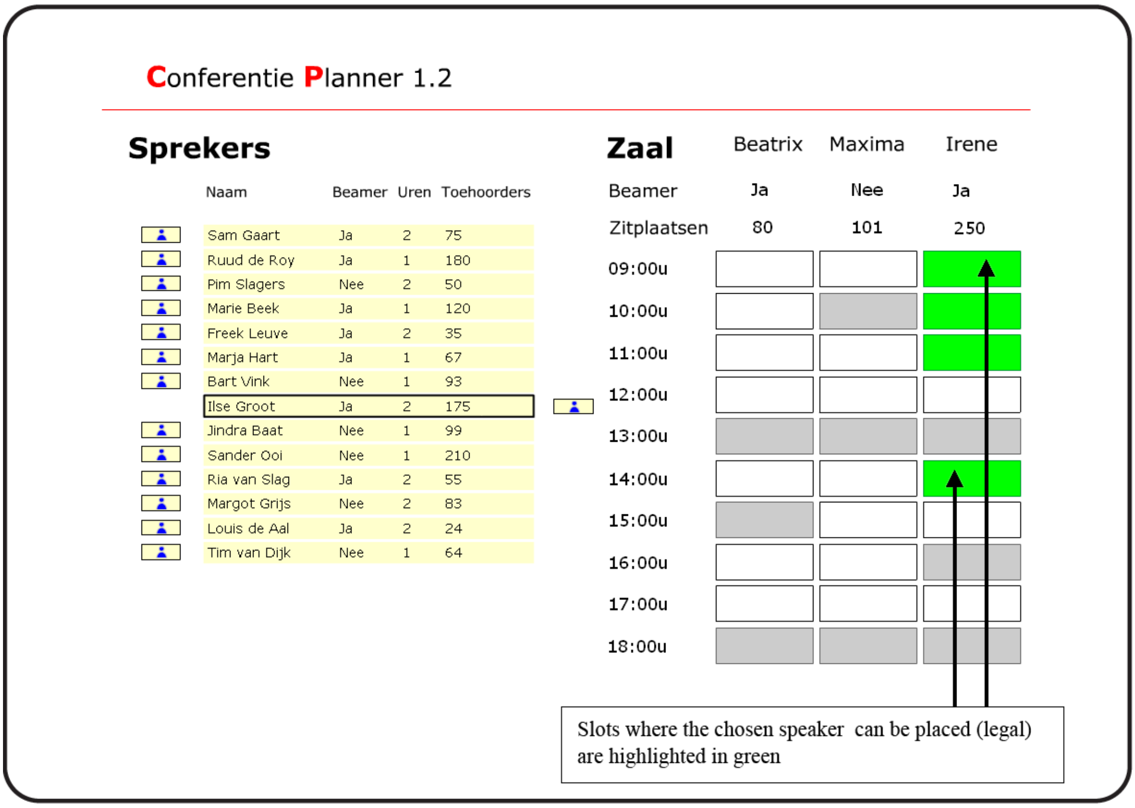



Figure 4.2 - Conference planner application: The Externalization interface

## 4.3 Material: The Need for Cognition scale

In experiment 2, we investigated the influence of externally induced motivation by attempting to persuade users into certain behavior by means of specific instruction. In this experiment, we include possible differences in task conception or problem solving attitude, focusing on *more stable* personal qualities of people that could influence behavior. We will estimate the cognitive style of users by measuring their "Need for Cognition" (NFC), a construct that measures the tendency of individuals to engage in and enjoy effortful cognitive tasks (Cacioppo & Petty, 1982). Cacioppo and Petty (1982) did several studies, in which they tried to create such an instrument, resulting in the first NFC scale, which consisted of 34 items. They did a test for simple and complex information processing and need for cognition. In the experiment, subjects did a number-circling task with a table containing 3500 random numbers. In the *simple* number-circling task condition the subjects were instructed to circle all 1s, 5s, and 7s. In the *complex* number-circling condition, the subjects were instructed to circle all 3's, any 6 that preceded a 7, and every other 4. It showed that subjects with high NFC preferred the complex task, while subjects low on NFC liked the simple task better. It showed that NFC can be used to describe differences between (groups of) people that would otherwise remain concealed which is the main reason for including NFC in our current study. In how we adopt CLT, we constantly focus on how to increase Germane Cognitive Load which is associated with motivation and interest. It can be worthwhile to examine whether individual differences are at play, next to the effects that the two interface styles we use. Perhaps people whose intrinsic motivation is high already and who like to engage in, and enjoy effortful cognitive tasks are less sensitive to our interface style manipulations. To increase the efficiency of the assessment of the original NFC construct, Cacioppo, Petty and Kao (1984) developed a short version based on Cronbach's alpha, which we will use here. This version contains only 18 items and is therefore better usable in experimental settings. They explain that both persons high and low on NFC need to make sense of their world, but tend to do this in different ways. One could say that a person with a high NFC loves to seek, reflect on, and reason about information, whereas someone on the other end of the continuum only thinks as hard as necessary, and is inclined to rely on others. As a consequence, someone high on NFC has a more positive attitude towards problem solving or reasoning tasks compared to someone with a low NFC (Cacioppo, Petty, Feinstein, & Jarvis, 1996). NFC has consistently explained variations in individuals' predispositions to engage in varying levels of cognitive activity, making it reasonable to presume an effect on planfulness. The fact that high-NFC persons generally recall more information than low-NFC persons could have implications for HCI. Kort, Reilly and Picard (2001) emphasize the non-existence of a "one-size-fits-all" solution and the importance of adapting the external representation to the emotional state of the learner. In the same vein, literature on learning systems proposes adapting to user characteristics (e.g. Brusilovsky, 2001). NFC could possibly be used as a variable to adapt an interface to a specific user. Crystal and Kalyanaraman (2005) examined this possibility using a web information-seeking task. In their experiment, different versions of a website containing health information were used, in order to study whether NFC could influence the cognitive and emotional state of mind of participants. They compared clear vs. unclear link-labels and informative feedback vs. no feedback while pages where loading. NFC indeed influenced aspects of usability and user responses. Affective measures such as perceived speed and perceived ease of information location were positively associated with NFC. Regarding cognitive effects, it showed that NFC was negatively associated with information-seeking time and navigational moves; subjects that had a high NFC-score needed less time and made less moves. The experiment showed that NFC can indeed influence perceived system usability and the response of participants.

In the terminology of experiment 1 and 2, all this could mean that a person with a high NFC placed in the Externalization condition might still exercise a reasonably high amount of planning, as opposed to a low NFC subject, who is reluctant to plan. To see if any such interactions occur, it is therefore worthwhile to include this variable in our research. We use a Dutch translation of the 18-item NFC scale (3) implemented with Macromedia Authorware. Subjects have to rate statements on a Likert-scale (range 1-7), such as "*It's enough for me that something gets the job done; I don't care how or why it works*". Score "1" meant "strongly disagree" and score "7" meant "strongly agree".

## 4.4 Experiment 3: Conference Planner task, interface style and cognitive style

Subjects had to solve five tasks consisting of different conference scheduling situations using Conference Planner. The subjects were randomly divided over the two interface versions. There were always 14 speakers that had to be scheduled, and a solution always existed in such a way that all the before empty slots are filled with speakers.

### 4.4.1 Subjects

43 subjects recruited at Utrecht University participated in our experiment, 17 male, 26 female. Participants were 19 to 32 years old, and were following or had recently followed higher education. The experiment took at most one hour and subjects received a €5 reward.

### 4.4.2 Design

The experiment had a 2X2 design with two independent variables: interface style (Internalization or Externalization) and cognitive style (low or high NFC, based on the median). After the NFC-test, subjects were randomly assigned to one of the two interface versions. After the first 20 subjects, the NFC median was calculated. On this basis we labeled subjects High-NFC or Low-NFC. We continuously checked whether low/high NFC subjects were evenly divided over INT/EXT and from then on NFC was checked immediately (via an automated process) and the condition interface style was continuously re-balanced to obtain equal groups. At the end the subjects per condition were:

Table 4.1 – Design experiment 3 and number of subjects per condition

| | Low-NFC | High-NFC |
|---|---|---|
| Internalization | 11 | 11 |
| Externalization | 10 | 11 |

### 4.4.3 Measures

In each of the three parts of the experiment, we measured various dependent variables.

**NFC-questionnaire**

NFC-score: the average of the 18 answers (range 1-7). We used the Dutch translation of the NFC-test (Pieters, Verplanken & Modde, 1987).

**Conference Planner: Time-based and move-based measures**

- *Total time*: The average time needed to solve the tasks
- *Time before first move*: The time between the moment the problem appears on-screen and the first move. It is an indicator for planning, telling how long subjects analyzed the problem before they started working on it
- *Time between moves*: The time that passes between having placed a speaker, and picking up the next. We interpret this measure as a planning indicator
- *Duration of a move*: The time that passes between picking up a speaker and placing it, the time that the speaker is floating across the screen
- *Superfluous moves*: The problems have a shortest path solution, with an optimal amount of moves (speakers dragged from left to right) to solve them. E.g. in the case of a list with 15 speakers, the optimal solution would be 15 moves (placing all speakers correctly the first time). Superfluous moves are *all* the extra unnecessary actions besides this shortest path. We use this measure as the main performance measure, because it reflects the efficiency with which the task has been solved. This measure includes also

*correction moves*: all extra performed moves (more than the shortest path) that were made to "fix" non-optimal placements (a situation where one was stuck). Also *reconsidered moves* are among the superfluous moves: when subjects pick up speakers and while dragging change their minds and put the speaker back before placement on the grid

**Conference Planner: Strategy Analysis**

We looked (per task) at whether or not subjects started solving the problem with a smart strategy, by first moving the speakers that had the most stringent constraints. We analyzed all the video recordings of the subjects' trials and gave points (one or zero) on two variables:

- whether subjects took a *sequential* approach when placing the speakers (starting with the 1st speaker in the list and work his way down)
- whether subjects started with the speakers that had the *most stringent constraints*.

In our analysis, we looked at the total scores for the five trials, which imply a minimum score of zero and a maximum of five (for both variables). It was expected that Externalization subjects would take a sequential approach more often compared to Internalization subjects, because subjects who are required to internalize plan more and therefore select the most appropriate speaker to be placed. Subjects in the Externalization condition on the other hand select a speaker without planning, probably resulting in a sequential approach. In the same vein, the subjects who are required to internalize are expected to plan according to the "most constraints first" strategy more often than Externalization subjects.

**Post Experimental Questionnaire**

The measures obtained from the post experimental questionnaire consist of answers to questions that tested for (a) declarative knowledge, (b) procedural knowledge, and (c) subjects' opinions. For each essay question, a subject could receive a maximum of two points. Most of the knowledge questions explicitly asked for an explanation ("why?"), besides a "yes" or "no" answer. A picture of a situation as in the application they just worked with accompanied most of the essay questions. The opinion measures indicate how the tasks were perceived by subjects themselves. Subjects could rate various statements on a scale from one to seven. The questionnaire consisted of three sets of questions:

- *Declarative knowledge*: Essay questions about *situations* that could be legal or illegal were shown, and one had to decide whether the situation was possible (could it occur regarding the constraints) and why (not).
- *Procedural knowledge:* Essay questions concerning solution procedures. Parts of situations were given and the questions were formulated as "how would you do this?" and "do you think this is a smart thing to do?"
- *Opinions:* Questions about opinions of subjects. There were 8 questions concerning perceived own problem solving and opinions about the application.

### 4.4.4 Hypotheses

**H1: Internalization leads to a more plan-based strategy and better performance than Externalization.**

Having information externalized, providing on-screen assistance tempts users to rely on the interface and *not* to form plans. The internalized condition lacks this guidance, encouraging planning and thinking before acting. We therefore expect that requiring subjects to internalize the information will lead to more elaborate planning and consequently better task performance such as more efficient solutions. We also expect better knowledge afterwards.

**H2: High NFC leads to a more plan-based strategy than low NFC, particularly in the Externalization condition**

We expect an interaction effect of NFC and interface style; interface style will have different effects, depending on whether a subject's NFC is high or low. We expect that subjects who already have high NFC scores will not perform so dramatically different between the two interface styles, since they like this kind of cognitive activities and already have an intrinsic motivation to plan their actions. With the Low NFC subjects however, we expect greater differences between the two interface styles. Low NFC subjects already have little internal motivation to start thinking about the task, and when they work with the Externalization interface, which seduces them to use a display-based approach and reason even less by themselves, we expect these people to show the least plan-based behavior and perform worst of all. We expect subjects that have low NFC, but work in the internalized condition to perform better than the former group because the interface seduces them to use a plan-based approach. In general, we also expect subjects with high NFC to perform better than low-NFC subjects.

### 4.4.5 Procedure

Subjects start with filling out the NFC questionnaire, followed by a textual introduction and a small video fragment showing how the application and its controls worked. After this, subjects start to work on the 5 problem solving tasks. Finally, the electronic post-experimental questionnaire has to be completed. After completing the five trials of the conference-planning task, subjects are asked to complete a final questionnaire (implemented with Macromedia Authorware), measuring knowledge of the problem and subjects' opinions.

### 4.4.6 Results Experiment 3

We analyzed the effects of interface style and cognitive style using ANOVA. Results with p-values between .05 and .10 are reported as tendencies. All the tasks were eventually solved correctly by all the subjects across conditions. We only report significant effects ($p < .05$) and tendencies ($p < .10$).

**The Need for Cognition scale**

The subjects scored a minimum of 2.39 and a maximum of 6.50 on the NFC scale. The mean score of the participants was 4.89 (SD 0.83), with the median at 5.06. Cronbach's alpha coefficient of the 18 statements was 0.89. High or low NFC had no effect, nor interaction effect on any measure, therefore we will not mention it anymore in the reported results henceforth. We will limit the reported data to differences between the Internalization and Externalization interface style.

**Time and move based measures per interface style**
In the table below are the different time and move-based measures per interface style.

Table 4.2 - Time and move based measures per interface style

| Average for the 5 tasks | Internalization | | Externalization | |
|---|---|---|---|---|
| | Mean | SD | Mean | SD |
| Total time needed (s) | 139.7 | 34.3 | 132.7 | 33.4 |
| Time before first move (s) | 19.8 | 7.4 | 15.3 | 7.2 |
| Time between moves (s) | 4.8 | 1.4 | 3.9 | 1.3 |
| Duration of moves (s) | 2.0 | 0.5 | 2.2 | 0.6 |
| Superfluous moves | 2.5 | 2.5 | 4.3 | 3.1 |

**Total time: time needed to complete a task**
The average total time that Internalization and Externalization subjects needed to complete the tasks did not differ between the two interface styles.

**Time before first move**
There was a significant main effect of interface version on the time that passed *before* subjects made their first move $F(1,41) = 4.18$, $p < .05$. Internalization subjects took more time before starting working on the task than Externalization subjects, $M = 19.8$, $SD = 7.4$ vs. $M = 15.3$, $SD = 7.2$.

**Time between moves**
There was a significant main effect of interface style on the average time *between* moves $F(1,41) = 4.79$, $p < .05$. Internalization subjects took more time between moves (in seconds), $M = 4.8$, $SD = 1.4$ vs. $M = 3.9$, $SD = 1.3$.

**Duration of a move**
The duration of the moves (time that passed between picking up a speaker and placing him) did not differ between the two interface styles.

**Superfluous moves**
There was a significant main effect of interface style on the number of superfluous moves that were made $F(1,41) = 4.37$, p <. 05. Internalization subjects made fewer superfluous moves than Externalization subjects, $M = 2.5$, $SD = 2.5$ vs. $M = 4.3$, $SD = 3.1$.

**Strategy analysis**
Internalization subjects showed a tendency to use the “most constraints first” strategy more often than Externalization subjects ($F(1,41) = 3.21$, $p < .08$). Internalization subjects used that strategy 2.4 times ($SD = 1.76$) out of 5 (tasks) whereas Externalization subjects used it 1.5 times out of 5 ($SD = 1.5$). Regarding using the sequential approach, for this measure there was no significant difference between the two interface styles. We conclude that the

Internalization condition used the "most constraint first", and that the Externalization subjects picked their speakers more randomly, not based on a specific sequential strategy.

**Results from the post experimental questionnaire**

*Declarative knowledge*

The effect of interface style on answers to the declarative knowledge questions was almost significant, at $F(1,41) = 3.73$, $p < .06$. Internalization subjects answered more of those questions correctly than Externalization subjects did ($M = 8.0$, $SD = .2$ vs. $M = 7.7$, $SD = .7$).

*Procedural knowledge*

Neither interface style nor cognitive style had an effect on procedural knowledge.

**Opinions**

On all of the questions Internalization subjects scored marginally higher, but only one of these differences was significant, namely whether the subjects "sometimes did not know how to proceed with the arrangement of the speakers", $F(1,41) = 5.91$, $p <. 05$. Externalization subjects scored significantly lower ($M = 5.24$, $SD = 1.61$) than subjects in the Internalization condition ($M = 6.23$, $SD = .97$), meaning that they had this feeling more often.

### 4.4.7 Conclusion Experiment 3

This experiment investigated the influences that interface style and cognitive style have on planful behavior from the user and consequently on problem solving performance. As in earlier experiments, it once again showed that also in this more realistic task user behavior differed depending on interface style (see also Van Nimwegen & Burgos, 2005; Van Nimwegen, Van Oostendorp, Burgos & Koper, 2006; Van Nimwegen, Burgos, Van Oostendorp & Tabachneck-Schijf, 2006).

Let us start by broadly answering our specific research questions of section 1.5.1 that can be addressed with this experiment. Question 1 concernes repeated problem solving performance between the two interface styles, and it showed on several measures that Internalization subjects were at advantage. They thought longer before they started and between individual moves, which indicates planfulness and mental effort. They came up with more direct solution paths, apart from the indication that their approach (strategy) tended to be better. The fifth question concerned the attitude individual users have to problem solving in general. We conclude that this did not influence our performance measures in any way, it was the interface versions alone that account for the found differences.

The first hypothesis (section 4.4.4), stating that the Internalization version of the task leads to a more plan-based strategy and better performance than Externalization, was supported from various measures. Having information externalized, providing on-screen assistance tempted users to rely on the interface and *not* to form plans.

The total time needed for the tasks was the same between the conditions, but Internalization subjects took more time before starting each task, and also took more time between placing one speaker and the next. This indicates longer thinking time, deeper processing from the user. This is in line with results of O'Hara and Payne (1998), who reported a longer inter-move latency for subjects in the effortful condition, indicating a more plan-based approach. Then the question might rise why there was no difference in overall time, regarding the fact that Internalization took more time on the two mentioned measures. This can be explained with the superfluous moves measure, on which Externalization subjects had a higher score in general. Making extra superfluous moves also costs time, which in the end caused that there were no overall time differences.

Scrutinizing on these superfluous moves, it has to be born in mind that the issue here was not "can they solve it?" but "how smart or efficiently do they solve it?", since in the end each problem had a solution, and these tasks are not extremely difficult. Internalization subjects solved the problems with fewer superfluous moves, in a more straightforward manner, thus with less deviation from the minimum amount of moves, resulting in greater efficiency. We infer that these more direct paths in the problem space are a result stemming from better planning. The superfluous moves included "correction moves", moves made to "repair" situations that sometimes were created by users making "not so smart" moves. Superfluous moves also included "reconsidered moves" (when a user started a move but changed his mind and puts it back, like a chess player who realizes his mistake a fraction after he picks up a piece).

Regarding the strategy that subjects chose to use, also a qualitative analysis of the results pointed at a more plan-based approach by the Internalization subjects, showing that those subjects more often filled the timetable by first scheduling speakers with the most constraints. This strategy is again an indicator of good planning, because it shows that people think about whom, and how they are going to schedule *before* starting with the task. We will elaborate more on the issue of strategy use in chapter 5. We also tested the knowledge subjects possessed after completing the tasks. In the declarative knowledge questions, subjects had to judge situations where the rules sometimes were violated. They had to identify whether the shown situations were theoretically possible or not. They had to look at the constraints of a situation, and decide whether any "rule" was violated (such as a speaker with an audience of 120 placed in a room where only 75 fitted). The expected effect of interface style on declarative knowledge questions was almost significant, although not as strong as in experiment 1 and 2 with Balls & Boxes. It showed that Internalization subjects tended to give more correct answers to those questions, which might be of value in for example a transfer task, or when doing a similar task after a considerable delay. An explanation for the better answers by Internalization subjects could be that they apparently spend the extra time (before, and between moves) on planning and thinking in all the preceding tasks. This explanation also gets some support from the above mentioned tendency of Internalization subjects to use a more appropriate strategy. There were also several procedural knowledge questions. These were not so much about judgment, but about insight on what to do to solve a part of a problem. Interface style had no effect here, no differences concerning answers to the procedural questions were found, since the problems were not so difficult and all subjects correctly solved al the situations. This pattern is similar to the one found in previous experiments: a positive effect of Internalization on declarative knowledge, and no effect on procedural knowledge. Lastly, the opinion questions indicated that all the subjects had little difficulty performing the tasks and were confident about their own performance. There were hardly any differences between the two interface styles, although the scores of Internalization subjects were marginally higher. One question yielded significantly different scores, pointing at an advantage of Internalization. We asked whether the subjects sometimes did not know how to proceed with the arrangement of the speakers. This “feeling stuck” occurred significantly more often in the Externalization condition than in the Internalization condition. This corresponds with the finding that Externalization subjects needed more superfluous moves (they had to make corrections more often) to complete the tasks. We take this also as an indication that the Internalization subjects had better insight.

The second hypothesis has to be rejected. Preexisting cognitive styles of users, at least on the dimension of high NFC versus low NFC had no influence; pre-existing attitudes towards problem solving did not have significant main effects or interaction effects on the displayed behavior and performance of subjects. Perhaps the effects of the interface style overruled possible effects of preexisting individual differences in cognitive style. However, the lack of findings may also be due to our subjects’ high NFC scores in general. The difference in scores between high-NFC and low-NFC may not have been large enough to produce significant effects. Another possible reason for the lack of findings is that our task might have been too easy for the subjects, not needing the extra impetus from having high NFC scores. This is supported by answers in the post experimental questionnaire concerning the perceived difficulty. Subjects reported having little difficulty with the tasks, and the cognitive effort required was in general quite low. With regard to CLT, the difference between high and low NFC does not seem to sufficiently influence Germane Cognitive Load (which must be as high as possible when solving a problem). The interface style is the deciding factor in provoking user behavior, not the preexisting attitudes as measured by NFC.

To sum up our conclusions, at not any point or with any measure did Externalization result in better performance, reconfirming various earlier findings. Also in a more realistic task, we found positive effects of *Internalization* on problem solving behavior: having the user internalize the information leads to a more plan-based behavior, smarter solution paths, better declarative knowledge, and less feeling lost. This presumed better problem solving behavior was also reflected in the strategy subjects chose, although this effect was justly not significant. Internalization subjects more often applied a strategy where scheduling started by identifying and placing the speakers with the most stringent constraints. Externalization on the other hand led to a more display-based approach, resulting in less efficient solutions (more moves) and less thinking ahead about moves to be made. It is worthwhile to reflect on what was in effect externalized and visualized. In the Externalization condition, this was the result of applying the set of constraints of speakers to the sets of constraints of the scheduling slots. The interface showed which actions are allowed with the object at hand. Subjects were able to see the outcome of the *application* of the rules, which refers to what we called destination feedback. This is comparable to the graying out of menu items in for example Microsoft Word, showing that the current situation does not permit those functions to be used. We showed that this widely accepted and implemented guideline does sometimes have undesirable effects. Regarding CLT, externalizing the application of rules, which can be thought of as assistance had negative effects and was not beneficial for contemplation (see also Schnotz & Rasch, 2005) and planning. We consider it to cause too high (unnecessary) Extraneous Cognitive Load. ECL being too high permitted less than desirable Germane Cognitive Load, and consequently too little cognitive resources are devoted to meaningful cognitive processes.

How can we link all this to general aspects of usability such as learnability, satisfaction, errors/memorability and efficiency? In previous research we saw that errors/memorability (especially memorability) was better in the Internalization condition (section 3.3.6 and 3.3.7). Concerning errors, the current results concerning superfluous moves (although they are not the same as errors) point at an advantage for Internalization. Efficiency in terms of “path economy” was better in the situations where subjects had to internalize information themselves. Satisfaction was not our main focus, but there was a question reflecting satisfaction, and there was no difference between the interfaces. In short, besides for user satisfaction on which we did not find differences, when sticking to the main usability guidelines as mentioned in section 1.1, the “unfriendly” versions of our applications, resulted in higher usability, although the common guideline to externalize information was severely violated.

In the next section we will study transfer, and at the issue of interruption as well. This can be interesting in interaction paradigms in a nowadays-prevalent area: mobile devices. Our devices are more mobile than ever before and during interactions with these devices a wide range of interruptions and distractions are commonplace. Typical interruptions of our modern lives are getting in/out of a taxi, reaching a subway station or temporary losing your wireless connection (Nagata, 2006). One could argue that it is preferable that users resume interacting on the basis of solid internalized plans and knowledge, rather than catching up on the task on the basis of an interface assisting the user to a high degree. If the former is the case, it means that users still need to construct solid plans.

## 4.5 Experiment 4: Interface style, interruption and transfer

In this experiment, we will focus on two different aspects that are part of our research questions, namely interruptions and transfer of skill.

### 4.5.1 Transfer

Two tasks can be related to one another by sharing underlying common principles: the rules we learn in doing one task may be extended to perform some other task. Transfer occurs in a specific situation as a result of experience from another situation. When problem solvers apply previously acquired knowledge, transfer can occur in the new context.

In our research questions, there were two questions regarding transfer of skill. Question 2 focuses on whether learning and doing a task in one interface style influences performance on a similar (but different) task, and question 3 investigated whether learning and doing tasks in one interface style influences performance on that same task with another interface style. We ponder that the Internalization interface fosters transfer of skill more the Externalization interface. The idea behind is that Externalization hinders performance, because no strong mental plan or procedural trace has been formed, providing no or little flexibility. *If* however deeper levels of processing have occurred, instigated by Internalization, this could result in an advantage on a transfer task. In aforementioned research by O’Hara and Payne (1998, 1999) also transfer was studied. It showed that too strong a reliance on external information led to negative effects concerning planning and transfer of skills. Task knowledge acquired with the high-cost interface could more easily be transferred to problems within the same domain. The authors warn that just freeing up cognitive resources is not enough. The design of the problem solving environment has to encourage the use of these resources for planning as well. “To encourage the use of resources for planning” is equivalent with triggering Germane Cognitive Load as CLT states it. For the latter to happen, according to CLT, the design of the environment (designers can have control over this) should not cause too high Extraneous Cognitive Load. In short, strong reliance on the visual display may result in less planning and less transfer of skill. In experiment 2 we briefly looked at a transfer task that had exactly the same underlying structure, but looked and felt quite different at the surface. The Internalization subjects had a better performance (they solved it faster the first time and they solved it more often within the same time span) than the Externalization subjects. This suggests that Internalization caused more solid encoding and imprinting of strategies. In this experiment we will look at it more closely and systematically.

**Switching interface**

Does learning and doing tasks in one interface style have an influence on performing that same task afterwards with the other interface style? We will investigate what happens when subjects switch from EXT > INT. We expect that if one has learned and performed a task with an interface style that *has externalized* information, doing that same task with the interface that *lacks* Externalization (Internalization) will cause deteriorated performance. This is because users first partially relied on the guidance by the Externalization interface, and their internalized knowledge will be suboptimal. When this guidance is suddenly missing, they will perform worse.

We will also examine switching the interface the other way around, having subjects switch from INT > EXT. We expect the performance to be at the same level since users are already used to do the task based on information processing performed in their own memory.

**Transfer of skill to another task**

Does learning and doing a task in one interface style have an influence on performance on a similar (but still different) task? We expect that if one has learned and performed a task with an interface style that lacks externalized information (requiring a lot of Internalization from users) this allows high germane CL, which in turn fosters metacognitive activities such as planning. This in turn will result in better strategies and better imprinting of task knowledge. We expect users that worked with the Internalization interface are better able to transfer and apply this knowledge to different, related tasks.

### 4.5.2 Interruption

If users have learned a task by doing it several times with an interface where they had to internalize information themselves, we expect better command of, and better imprinted knowledge about the task, rules and solution procedures. We will impose an interruption of several minutes that is severely disrupting on the users in both the Internalization and the Externalization condition. Our assumption is that if actions are more planned, and if indeed knowledge and procedures are better represented in the Internalization condition, this will result in smoother and more efficient task resumption for users that had to internalize all the information by themselves. The 8-month interval between experiment 1 session A and B can hardly be called an interruption, but retesting of the same persons resulted in an advantage for the Internalization subjects. However, this intermezzo did not come as unexpectedly as interruptions mostly do, but was announced and took place *after* a task or a long interval, not at an inopportune moment, making it less disrupting. Nevertheless, we take the results as indication that Internalization indeed leads to more solid encoding and imprinting of strategies.

In the way we connect our research issues to CLT, the situation where a user has to internalize information him/herself was positioned on the Extraneous Cognitive Load continuum as "low" since little unnecessary processing is taking place, which permits high Germane Cognitive Load, devoted to processing, construction and automation of schemata. This in turn results in more efficient and more planful behavior, because an adequate portion of Working Memory *can* be devoted to effort involved in processing and construction of schemas or plans. According to CLT, high GCL leads to better encoded schemas. This corresponds to how we expect Internalization to lead to more plan-based behavior and more processing effort. If task resumption after a severe interruption is smooth, one might assume that the problem was tackled with some form of plan. Consequently, we expect Internalization to "survive" interruptions better. When task resumption poses a problem in the sense that a subject does not really know anymore where he was (as for example can be indicated by making lots of errors after the interruption), one might assume that problem solving did not really take place in a plan-based manner, but more in a display-based manner (Externalization).

Regarding interruptions, it is worthwhile to mention research by Oulasvirta and Saariluoma (2006). They connect their work on interruptions to the theory of long-term working memory (LTWM) versus short-term working memory (STWM) (Ericsson & Kintsch, 1995). This theory sees LTWM as an intermediate memory, which is part of "long term memory". Information can be encoded in LTWM into organized retrieval structures. STWM is a capacity and time limited store that updates and manipulates representations, and can contain pointers to structures in LTWM. Oulasvirta and Saariluoma (2006) state that if one is to survive an interruption, representations of the task must be stored so that they can be later accessed rapidly and in an error-free manner. They call this outcome of cognitive task processing "safeguarding of task representations". A suggestion for interface design to support interruption tolerance is to actively support and facilitate the development of memory skills. This is in accordance with the beneficial effects we expect from requiring to internalize information and consequently allowing high GCL.

Our interruption will be severe and occurs at an inopportune moment as happens in real working situations. We mimic a power failure, which besides interrupting is also upsetting: the power went of, and the computer screen turns black. Also the experimenter "suffers" this power failure and acts surprised, to make it more credible. Our expectation is that users of the Externalization version are affected by the interruption more because they work in a shallower manner. The Internalization group is less (or not) affected by the interruption; they work on the basis of a more solid and quickly accessible plan.

### 4.5.3 Material: Initial phase

In the initial phase uses the same Conference Planner application as in experiment 3. The only difference here is that all the trials had 14 speakers (in experiment 3 this was varied). This was done to be better able to compare between tasks without having to resort to z-scores. The interruption, the mimicked partial power failure in the room was initiated by the experimenter who monitored the process out of sight, and pushed a switch after the 4th speaker was placed.

### 4.5.4 Material: The transfer phase (Ferry Planner)

After the five trials with the Conference Planner application, the subjects had to solve 2 transfer tasks with another application (Ferry Planner), in which coaches had to be fit on a ferry deck. Although in essence it is the same problem as Conference Planner, it looked and felt different, and the orientation was not right-left as before, but up-down (Figure 4.3). However, the underlying task structure was the same as in the Conference Planner, making it a “near transfer” (Perkins & Salomon, 1992). The constraints of conference planner were translated to “width of the coach”, “weight” and “coach length”. Externalization was implemented in exactly the same way as in the Conference Planner.

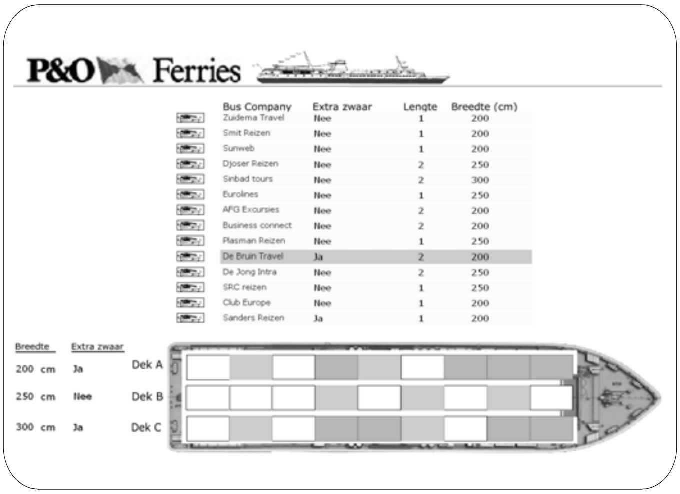



Figure 4.3 - The Ferry Planner in the Externalization version: legal timeslots turned green (in this image: dark grey)

### 4.5.5 Subjects and Design

The 45 participants were all students from Utrecht University and received a €6 reward. The dependent variables we focus on are time and path measures. Subjects receive seven constraint-satisfaction planning tasks in two phases (5 in the initial phase and 2 in the transfer phase). In the initial phase, the subjects will be randomly divided in 2 groups (Figure 4.4). In the initial phase, we look at differences in performance *between the two interface styles* on 5 trials with the Conference Planner application. Until task 3, the course of events is the same as in the previous experiment, but in task 4, a severe interruption (an electrical power failure) will occur to see which group is more affected by this. After this, a fifth “normal” task follows (no more interruptions).

In the transfer phase, all the subjects receive the *transfer application* in task 6, the application changes to Ferry Planner, which shares the same underlying principles. Half of the subjects, group B and C switches *interface* during this 1$^{st}$ transfer task (task 6), the other half does so later on in task 7 (group A and D). This allows us to investigate the effect of the *order* of the interface switch (INT→EXT or EXT→INT) and whether the performance depends on *when* it takes place (in the 1st transfer task, or 2nd transfer task when subjects already worked with the application). We can also compare the influence of doing the 1st transfer task with *another* interface (INT→EXT, EXT→INT, groups B and C) than before, vs. doing it with the *same* interface version as before (INT→INT, EXT→EXT, groups A and D).

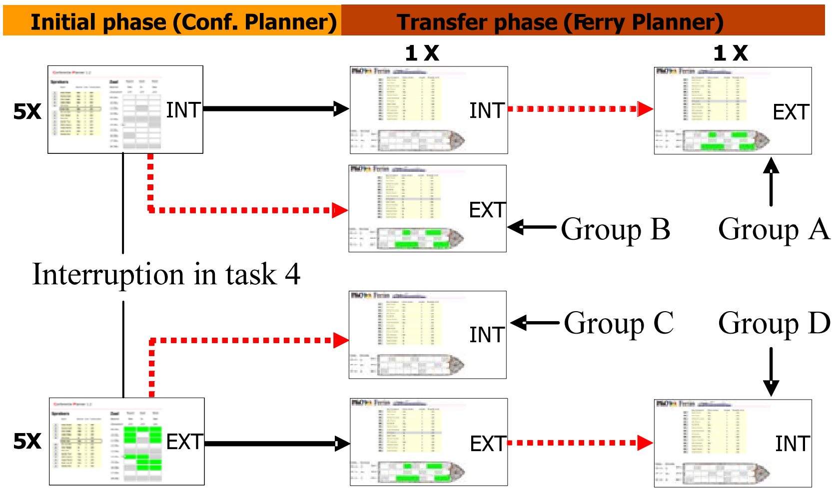



Figure 4.4 - Experimental design experiment 4

The independent variables in the 2 phases of this design are:

**In the initial phase**

- Interface version (Internalization vs. Externalization, the INT-EXT groups before they were divided into Group A, B, C and D)

**In the first transfer task**

- Whether the **first** transfer task is done with the same interface (group A & D) vs. with another interface (group B & C) than before

**Over the two transfer tasks**

- *When* the **interface** switch took place (group B & C in the 1st transfer task vs. group A & D in the 2nd transfer second task)

### 4.5.6 Procedure and instruction

The experiment took about 45 minutes. Subjects performed the tasks, and for all the subjects during the 4th task after placing the 5th speaker the system “crashed”. The monitor went off, making the situation look like a power failure in a certain electrical circuit. The room lighting, however stayed on, as did one PC other at the other side of the usability lab. Also the experimenter’s PC went off, and the experimenter played along and mimicked surprise:

*“Oh…my PC also went off; something seems to have gone wrong. I’ll look into it. For now, do the next task on another computer, while I try to fix the problem here”.*

Subjects were accompanied to a “working” PC out of sight, behind a screen on the other side of the room, which *did* work and were given a questionnaire (unrelated to the tasks). After 3 minutes, they were called back and were told:

*”I fixed the problem. The state at which it crashed has been recovered, so you can continue where you were”.*

### 4.5.7 Hypotheses

**H1: The effect of the interruption**

Our expectation is that users of the Externalization version are affected by the interruption more because they work in a shallower manner. The Internalization group is less (or not) affected by the interruption; they work on the basis of a more solid plan.

**H2: The effect of interface style in the initial phase on performance on the first transfer task**

Subjects who worked with the Internalization interface in the initial phase will perform better on a transfer task than subjects who worked with the Externalization interface. Internalization provokes better strategies in the initial phase. They solve problems with more efficient solution paths, reflected by fewer superfluous moves. Working in this manner in the initial phase will have a positive impact on how the transfer task will be approached.

**H3: The effect of order of switching interfaces**

Performance after switching interfaces in the sequence Internalization → Externalization (INT→EXT) will result in better performance than switching in the sequence Externalization→ Internalization (EXT→INT). Switching from "no guidance" to "guidance" causes no problem, since "no guidance" subjects already actively developed strategies to approach the problem. The subjects that switched from "guided" to "unguided" relied on guidance and will suddenly have to think more actively. They do not develop appropriate approaches yet, whereas Internalization subjects already did this.

### 4.5.8 Results Experiment 4

**General**

All the subjects correctly solved all seven tasks in the two phases. As in earlier research, we did not find overall differences in completion time between the conditions. However, as will see in the next section, the quality of solution paths during that same time span *was* different.

**Results of the initial phase in which the interruption took place**

We expected that the Internalization interface would lead to low ECL and high GCL. We consequently expected better motivation and higher effort, resulting in smarter and better imprinted strategies leading to solutions that are more efficient. Below are the results of a repeated measures analysis involving interface style, and the task (there were 5 tasks).

Table 4.3 – Time and move based measures per interface style

| Average for the 5 tasks | Internalization | | Externalization | |
|---|---|---|---|---|
| | Mean | SD | Mean | SD |
| Total time needed (s) | 141.3 | 7.7 | 143.1 | 46.8 |
| Time before first move (s) | 12.3 | 3.4 | 9.8 | 4.6 |
| Time between moves (s) | 5.5 | 1.3 | 3.6 | 1.3 |
| Duration of moves (s) | 2.3 | 0.4 | 2.7 | 0.6 |
| Superfluous moves | 2.9 | 1.5 | 8.7 | 8.5 |

**Time based measures per interface style**
We used the same measures as in experiment 3, and mostly the same patterns were found:

*Total time*
As in experiment 3, the average time needed per task did not differ between conditions ($p > .05$).

*Time before first move*
As in experiment 3, Internalization subjects took significantly more thinking time before moves than Externalization subjects, this was a main effect $F(1,43) = 4.28$, $p < .05$. The averages for Internalization and Externalization subjects, were $M = 12.3s$, $SD = 3.4$ and $M = 9.8s$, $SD = 4.6$ respectively over all the tasks.

*Time between moves*
Also the difference in time between move was significant, $F(1,43) = 23.3$, $p < .001$. On average, over the 5 tasks, Internalization subjects ($M = 5.5$, $SD = 1.3$) took more time between moves than Externalization subjects ($M = 3.6$, $SD = 1.3$).

*Duration of moves*
In contrast to experiment 3, this time there was also a significant main effect of interface style on the average time *between* moves, $F(1,43) = 5.5$, $p < .05$. On average over the 5 tasks, Internalization subjects thought longer between moves as seen above, but the actual moving of the speaker itself they did faster shorter ($M = 2.3$, $SD = 0.4$) than the Externalization subjects ($M = 2.7$, $SD = 0.6$).

*Superfluous moves*
Repeated measures ANOVA showed that (as before in experiment 3) there was a main effect of interface version on the mean number of superfluous moves, $F(1,42) = 8.63$, $p < .05$ per task. Subjects that worked with the Externalization version came up with less efficient solutions, making more superfluous moves than Internalization subjects ($M = 8.7$, $SD = 8.5$ vs. $M = 2.9$, $SD = 1.5$) on average. Furthermore, there was an interaction effect, $F(4,42) = 2.34$, $p < .05$. The interaction effect involves two things. Firstly, the difference in superfluous moves decreased fast. In Figure 4.5, it is visible how in the beginning the Externalization subjects perform many more actions than Internalization subjects do. In task 2, this difference is already smaller and in task 3 the difference is not significant anymore, Externalization subjects improved their performance from task 1 to task 3. The second issue concerns the main focus of the experiment, the influence of the *interruption* in the 4$^{th}$ task as stated in hypothesis 1. After placing the 5th speaker, the power of the PC went off.

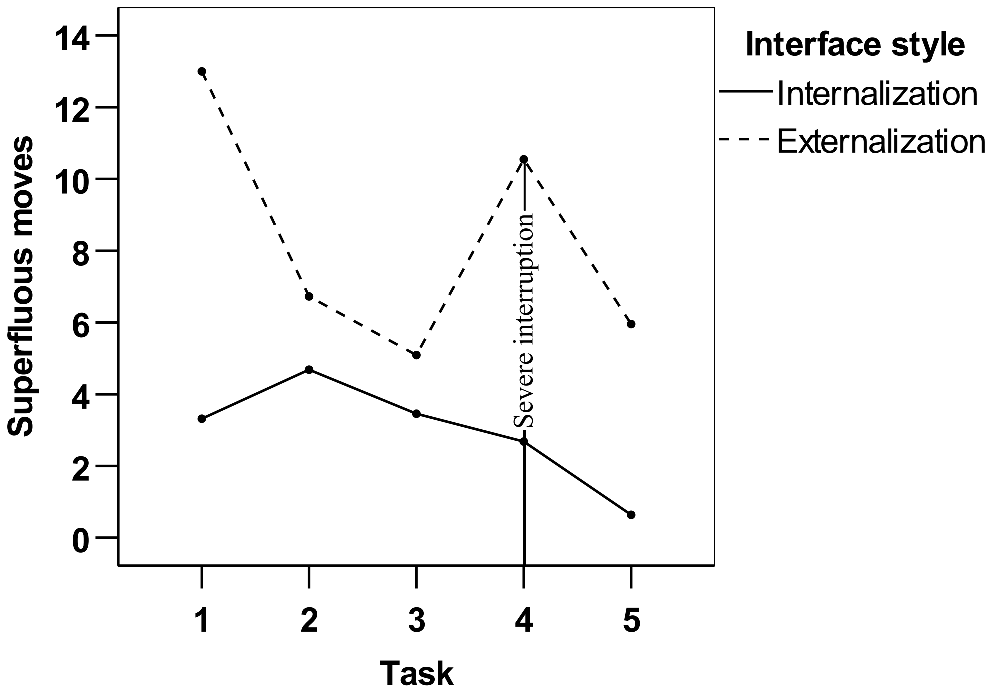



Figure 4.5 – Mean number of superfluous moves per task and interface version in the initial phase

Post hoc analysis revealed that the difference in superfluous moves between the two versions was significant for task 4 and 5. The interruption had a different effect on the Externalization subjects than on the Internalization subjects. On task resumption, Externalization subjects fell back dramatically, again, making many more superfluous moves than before, while Internalization subjects calmly went on, with fewer superfluous moves. We also asked subjects how they perceived their own efficiency. Internalization subjects rated their own efficiency higher than Externalization subjects on a 7-item Likert-scale $T(44) = 2.10$, $p < .05$. The Internalization version made people feel they performed the task better, as has been shown by the data already ($M = 5.2$, $SD = 1.1$ vs. $M = 4.4$, $SD = 1.3$).

**Results of the transfer phase**

*Performance on the first transfer task using the same interface as before vs. the other interface*

For all the subjects, task 6 was a different task. Some also switched also interface at the same time, others later on in task 7. To compare performance on the transfer task *itself*, we have to look at how well all subjects did on the *first* task of the Ferry Planner application, in task 6 when the task was still new for all the subjects. We thus look at doing the transfer task *with the same interface as in the initial phase* vs. doing the transfer task *with a new interface* by looking at the performance on the first transfer task for all the subjects (testing hypothesis 3). There was a main effect of interface version subjects worked with in the initial phase, $F(1,41) = 5.82$, $p < .05$, on superfluous moves on the first transfer task. In general, subjects who worked with Externalization before made more superfluous moves on the first transfer task than the ones that had Internalization before (M = 6.35, SD = 2.57 vs. M = 3.14, SD = 5.97). However, Figure 4.6 suggests an interaction effect. There was indeed a significant interaction of the interface subjects worked with before, and the one used in the transfer task, on performance in terms of superfluous moves, $F(1,41) = 4.22$, $p < .05$. Post hoc analysis revealed that performance on the transfer task between groups A and D (who used the same interface style as before) did not differ (INT→INT and EXT→EXT). Also, performance of group B that did the initial phase with the Internalization interface and then the transfer task with the Externalization interface (INT→EXT) did not significantly differ from groups A and D.

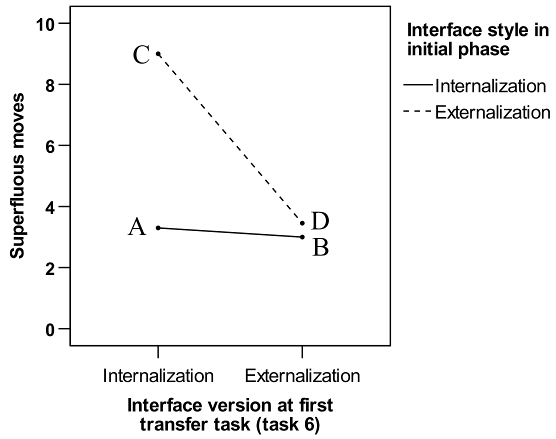



Figure 4.6 - Mean number of superfluous moves: Interface switch vs. no interface switch at transfer

However, while doing the transfer task, switching from Internalization to Externalization (from guidance to no guidance) did cause a much worse score on the transfer task. Group C had significantly more superfluous moves than the other 3 groups. In other words, doing the *transfer task* with the same type of interface as before, or using EXT after having used INT results in the same (low) amount of superfluous moves in the first transfer task. But using INT after EXT (as in Group C) caused a problem: the efficiency of solutions on the transfer task was significantly worse.

*The order in which the interface styles were presented*

Besides having another application (Ferry Planner) also the interface subjects worked with, was varied in the transfer phase. In the transfer phase, all the subjects changed interface version at some point, be it in task 6 (first encounter with the transfer task) or in task 7 (doing the transfer task for the second time). The red dashed arrows in figure 4.4 represent the switches. Half of subjects switched in the sequence INT→EXT (group A and B, figure 4.4). The other half of the subjects switched in the sequence EXT→ INT (group C and D). This allows us to investigate the influence of the order of switching, in combination with *when* this is done (testing hypothesis 2).

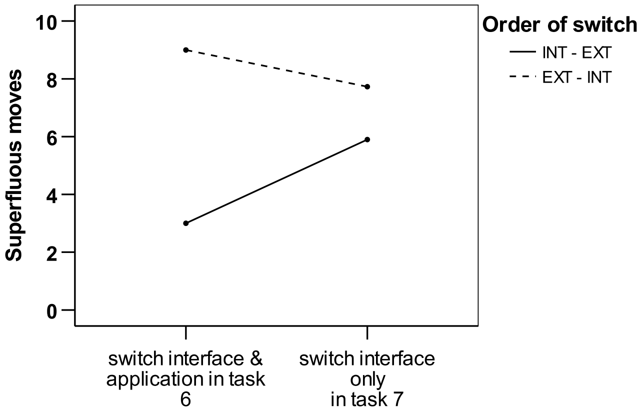



Figure 4.7 - Mean number of superfluous moves: Order of interface switch and when it took place

There was only a significant main effect of the *order of the switch*, $F(1,43) = 5.92$. $p < 0.05$ (Figure 4.7). Switching to the "new" interface in the sequence INT→EXT always resulted in fewer superfluous moves than in the sequence EXT→INT. Regardless of whether the switch took place in task 6 or 7, subjects that switched from INT to EXT solved the tasks more efficiently ($M = 4.43$, $SD = 4.1$ vs. $M = 8.39$, $SD = 6.74$).

### 4.5.9 Discussion Experiment 4

In general, as in the three earlier experiments we did not find any advantage of the interface that provided guidance (Externalization). In the initial phase (with the Conference Planner) Internalization subjects performed better right from the beginning. During the initial phase the performance of Internalization and Externalization subjects converged; Externalization subjects improved their performance by task 3, making fewer superfluous moves than before.

Hypothesis 1 can be accepted. The severe interruption that followed in task 4 affected the Externalization subjects negatively, whereas it had no influence on the performance of the Internalization subjects. Immediately after the interruption Externalization subjects fell back significantly, again making more superfluous moves. The fact that the Externalization subjects were more affected indicates that they had worse insight in how to resume the task. We interpret this as stemming from less solid plans and worse strategies and knowledge imprinted in memory (provoked by the guided interface). Interesting here is also the finding that subjects who worked with the Internalization version (INT) *correctly* judge their own efficiency higher.

The second hypothesis concerns performance in the *first* transfer task, when it was still new for all the subjects. We expected that subjects who worked with Internalization in the initial phase perform better on a transfer task than subjects who worked with Externalization. An assumption was that Internalization subjects solve the problems with more efficient solution paths. Having worked in this manner in the initial phase will have a positive impact on how the transfer task will be approached. This hypothesis is partly supported. An interaction effect of "interface in initial phase" and "interface in transfer phase" causes differences in performance on the first transfer task. Group A and B who both used the Internalization interface in the initial phase performed equally well, regardless of the interface version during transfer. However, contrary to our expectations, Group D who worked with Externalization before and had the same interface during transfer, performed equally well as group A and B. However, group C that did the initial phase with the Externalization interface and transfer in the internalized version performed significantly worse on the transfer task

Our third hypothesis concerned the order in which a switch can take place is also supported. Regarding transfer - the main focus of the current experiment next to the influence of interruption - we expected that subjects using the Internalization interface (provoking more proactive thinking) will not be affected by having to work with the Externalization interface. These former Internalization subjects will do the task in the other interface equally well, since they already master the underlying concepts and strategies. When the order is reversed (from EXT→INT), however, it is different, since here subjects are used to guidance and will suddenly have to do without it. We expected that working with the Externalization version first followed by working with the Internalization version would result in worse performance than the other way around. This was indeed the case. Overall, the subjects that had a series of tasks with the Internalization interface first performed better, and switching to the Externalization version did not change their performance. All along, the Internalization subjects solved the problems in a more proactive manner, without guidance by the system. The Externalization version only indicates legal moves, but when one is used to solve the problem entirely by oneself, it can be easy to ignore “new” feedback after switching interfaces. Internalization subjects kept working in the same manner, their strategies were better, resulting in more efficient solutions paths. Whether this took place between phases (transfer) or within the same phase, made no difference. As said, during the course of the initial phase all subjects solved the problems correctly eventually. The Externalization subjects had less efficient solutions in general, and are assumed to work more on a trial and error basis. When suddenly working with the Internalization version they probably continued to apply the trial and error strategies as they had done all the time.

## 4.6 General discussion Experiment 3 and 4

We broadly summarize the answer to the specific research questions that were addressed in experiment 3 and 4. We saw once more that in terms of efficiency, users who work with the Internalization version perform better (question 1). It showed that Internalization subjects outperformed Externalization subjects. They had better solutions, thought longer before they start on the problem, and longer between the separate moves (experiment 3). When switching from one task to another (question 2), it showed that working with the Externalization interface posed a risk, in that when an Internalization interface followed, performance decreased dramatically. Furthermore, we found that in general, when switching interface (question 3), the sequence Internalization – Externalization always resulted in better performance than the other way around. Individual cognitive style (question 5) was measured by the Need for Cognition scale, reflecting how people like or dislike to solve complex problems. This had no influence whatsoever; the interfaces themselves caused different performance. Lastly, a severe interruption (question 7) was survived better by Internalization subjects. Immediately after the interruption, Externalization subjects' performance fell back significantly. The fact that the Externalization subjects were more affected indicates that they had worse insight in how to resume the task and did not follow a very strong plan.

There was also the indication that the Internalization subjects showed a smarter strategy in approaching the problem, and that they judged their own efficiency higher. Lastly, Internalization subjects felt less lost, and had better knowledge after.

The fact that there were no overall time differences, combined with the efficiency measure "superfluous moves" is important. The Externalization subjects spent less time on thinking and planning and more time on trying out moves and making superfluous moves (making moves costs time). The Internalization group spends more time on planning and thinking, and less time on making (fewer) moves.

The above findings strengthen our assumption that deeper thought and more contemplation are a result of not guiding users too much and that this causes better performance in transfer situations. The results are in line with O'Hara and Payne (1998, 1999) who found that the subjects with a plan-based approach used fewer moves than subjects using a display-based approach. They also stated that "backtracking" (undoing a move and return to previous situation) occurred more in subjects showing display-based problem solving.

Are users of the Externalization interface so used to the information and feedback from the interface that they will be stuck after an interruption or in a new transfer situation? In this light, findings by Fu and Gray (2004) are relevant. They found that when inefficient procedures are chosen to solve a problem, these procedures share two characteristics. They are often generic and well-practiced, but more importantly the procedure is composed of interactive components that bring fast, incremental feedback on the external problem states. These actions require less cognitive effort, but this bias to depend on interactive units unfortunately leads to solution paths that require more effort in the long run.

With regard to CLT, can the presumed facilitating quality of externalized guidance as we implemented be seen as Extraneous Cognitive Load? Our findings strengthen our idea that Externalization as we implemented it does provoke behavior that is *not* directed at developing schemas and strategies. This is corresponding to what happens when Extraneous Cognitive Load is high, as mentioned in Cognitive Load Theory. This line of thought corresponds to the way we postulated the first of two scenarios in section 2.6. Furthermore, it can hold that Internalization resulting in low Extraneous Cognitive Load can correspond with high Germane Cognitive Load, allowing better motivation and mindful effort to look for what underlies the tasks. The latter line of thought corresponds with the way we postulated scenario 2 in section 2.6. Adopting CLT as we attempt, could be a promising in our research, because CLT sufficiently acknowledges issues such as effort and motivational factors (Sweller et al., 1998).

In experiment 4 the design was quite complex because of the difficult balancing between conditions, orders and versions. Group B and C solved only one transfer task, and group A and D solved two. To try to make transfer effects more clear, the issue of transfer will deserve more attention in the last experiment with conference planner in the next chapter. The tendency we found that Internalization provoked a smarter strategy made us decide to look closer into the issue of strategy. In order to examine whether subjects indeed are more oriented towards the construction of smart strategies and plans, we will formalize our measurement of strategy more in the next experiment. Furthermore, we continuously theorize about *where* users are looking and we mention “distractive” and “attention taking” features that we assume the Externalization version to have. By means of “eye tracking”, we will examine in the next experiment how actions are preceded by inspecting relevant areas of the screen where information is located. We use eye tracking data to determine the regions of the screen where users actually look at, but we also examine the patterns of gaze.

# 5 More focus on strategy and transfer and findings from eye tracking

In the former experiment, the design was such, that the subjects all received five tasks in the initial phase, and only one and sometimes two in the transfer phase. The experiment showed its value, in terms of the effect of the interruption, and there were also effects on transfer performance (Van Nimwegen & Van Oostendorp, 2007a, 2008a). In the fifth and last experiment that is described in this dissertation, we will scrutinize more on some of our previous issues.

Regarding the research questions in section 1.5, question 1 concerned how problem solving performance evolves over time, on various measures, depending on the interface style worked with. With eye tracking data we want to backup our conclusions regarding where, how often, and how long subjects fixate on certain regions. We also developed a mathematical measure indicating to which extent subjects are applying a smart strategy (a strategy with a high chance of an economical and direct solution) in their approach to solve the task. With a more formal strategy measure than before (see strategy analysis, section 4.4.6) we want to obtain extra evidence for our assumption that smarter strategies are constructed when working with the Internalization interface style. Lastly, we will again (but more profoundly) address question 2 concerning transfer of skill and whether doing a task in one interface style influences performance on a *similar task*. Question 3, also investigating transfer, but focusing on the influence of interface style on performing a task afterwards *with another interface style* will also be addressed. The design and balancing of conditions will be more straightforward and orthogonal than in experiment 4.

**More transfer**

As said, we want to focus on the transfer phase with more detail, because in experiment 4 possibilities were rather limited, since we could only say something about one transfer task (for all the subjects). Here we want to examine how performance evolves when more transfer tasks are performed. Although the transfer task is different, in essence, it is close to the initial task, and doing it several times in the transfer phase might result in a convergence of performance. This is what happened in experiment 4 in the initial phase where performance converged by task 3. We are curious to see what happens over time in the transfer task, so in the current experiment we will have the subjects do more transfer tasks.

**Strategy measure**

A problem solving strategy is a technique that may not guarantee solution, but serves as a guide in the problem solving process (Mayer, 1983). The way "guide" is mentioned here, is reminiscent of the term "plan-based", in that they both imply the working out of actions that have been thought over at an earlier point in time. In experiment 3 we had a brief look at strategy in the way of looking at whether subjects start with the most constraints first. It was a dichotomous measure which told us of each trial whether "most constraints first" was applied or not. We found that Internalization subjects tended to do this in more trials than the Externalization subjects did. In this experiment, we will use a more sensitive formalized mathematical measure, which classifies subjects according to their scheduling difficulty (based on their individual constraints). We thus want to be able to express to which extent the pattern of a "most constraints first" strategy has been applied by the subjects, with a continuous measure.

**Eye tracking**

We continuously theorized about *where* users are looking and we use terms as "distracting" and "attention taking". Regarding where users are looking, nowadays in usability studies often users' eye movements are tracked to determine what users are looking at or to discover patterns of interaction. However, some researchers have indicated that "mouse tracking" may be a low-cost alternative. Mouse movements are claimed to be linked to focus of attention of a user (Arroyo, Selker, & Wei, 2006). Chen, Anderson and Sohn (2003) found a correlation between eye and mouse movement, but they also note that it is *not always* possible to infer the user's attention from mouse cursor position or movement. This may be caused by the fact that users utilize the mouse in different ways. Some use it as a reading aid, while others move it to a blank area to get it out of the way or out of fear that they might accidentally click on something. We also conducted a minor pilot-study in which we wanted to see whether there is difference between mouse behaviors in *our* two conditions. Although this tool and the analyses that followed were interesting, it still did not tell us where users were looking (see Van Nimwegen & Van Oostendorp, 2006). To *really* know where one looks, eye tracking is indispensable.

We will attempt to backup our assumptions concerning "distractive" and "attention taking" properties (features that fall under Extraneous Cognitive Load) that we assume the Externalization version to have, with eye tracking data. A lot of attention is expected to be drawn to the green feedback in the grid in the Externalization condition (see Figure 5.9). We ponder that in the Externalization version subjects are to some extent "mislead" in the sense that they think that following interface cues is enough to reach a solution. This behavior is thought to discourage proactive contemplation. We will analyze whether there are relations between assumed cognitive activities and interface style, which are reflected in fixations and gaze patterns. We expect Externalization to cause more eye movements directed at the area on the right of the screen, where the (dynamic) destination feedback is taking place. Doing this is not so effective, and is an indicator for display-based behavior: take decisions on how to schedule speakers by looking which slots turn green and take a decision largely based on that. In the Internalization version, subjects do not have to process the externalized information, and the absence of this is thought to provoke more plan-based behavior, deeper thoughts and smarter strategies. This would in this case mean that they devote *more effort* to looking at constraints of speakers on the left side of the screen and think before picking up a speaker (we saw this happen already earlier in experiment 3 and 4). When the choice has been made, the speaker will be placed relatively fast on the scheduling grid, *less* time will be spend looking there.

## 5.1 Experiment 5: interface style, strategy, transfer and eye tracking data

### 5.1.1 Subjects and Design

Fifty-two subjects, mostly from the Linguistics department participated and received a €6 reward. In this experiment, in the initial phase half the subjects will be presented with 3 Conference Planner tasks with the Internalization interface, the other half with the Externalization interface. Then, after a 5 minute intermezzo in which they did another non-related task (the visual rotation task, see section 3.3.4) in the transfer phase (Ferry Planner) the two former groups were be split further (see Table 5.1). Of each group, half the subjects received the three transfer tasks in the Internalization interface and the other half the Externalization interface.

Table 5.1 – Design experiment 5 and number of subjects initial phase and transfer phase

| **Initial phase (3 x Conference Planner)** | **Transfer phase (3 x Ferry Planner)** |
|---|---|
| Internalization (n=26) | Internalization (n=13) |
| | Externalization (n=13) |
| Externalization (n=26) | Internalization (n=13) |
| | Externalization (n=13) |

### 5.1.2 Material: Conference Planner and Ferry planner

We use exactly the same software as in experiment 3 and 4, the Conference Planner and Ferry planner applications which both exist in an Internalization version and an Externalization version (with destination feedback).

### 5.1.3 Task between initial phase and transfer phase

Unlike in experiment 4, in order to establish a small pause between the two phases that will set the subject's mind to something completely different, we will use the exact same visual rotation task as described in section 3.3.4, during 5 minutes.

### 5.1.4 Strategy measure

In our tasks, arguably the smartest strategy with the highest chance of the most efficient solution is to start with speakers that have the most stringent constraints, since they need certain features, where others do not. Speakers varied on a combination of three variables: needed size of the room, needing a projector or not, or needing a 1 or 2 hour slot (Figure 4.1). One can of course make a typology of speakers by combining these variables, but a less complex way to look at it is to assign labels to speakers reflecting their "difficulty" in a simpler manner, corresponding with the nature of the grid where they have to be scheduled. There are always three rooms to schedule the speakers with their different constraints. Some speakers fit in all three the rooms, some in two of them, and the most difficult speakers fit only in one. Our measure "difficulty of speaker" has the values "1" (fits in only 1 room), "2" (fits in 2 rooms) and "3" (fits in all 3 rooms) where 1 is the most difficult type of speaker to schedule (lots of constraints) and 3 are the easiest speakers to schedule (no constraints, they fit anywhere). See Appendix D for an explanation of how these percentages were calculated. In the ideal situation, the speakers of difficulty 1 are placed at first, then the ones with difficulty 2, and at last the ones with difficulty 3 (they fit everywhere). From the log files of the subjects of each task, we can abstract the sequence in which speakers with varying difficulty were placed. For example in Figure 5.1, the course of events in a task where 14 speakers had to be placed is depicted. In this case, they were placed in 14 moves (thus no superfluous moves performed) and also in the ideal manner, reflecting a perfect strategy: starting with speakers of difficulty 1, and finishing with speakers that have difficulty 3. The 1-2-3 sequence, as we will call it henceforth, has been applied 100%. In Figure 5.2, however, the order of difficulty in the placed speakers seems quite random, and the 1-2-3 sequence is only moderately visible, 44.3%. This example represents a typical solution pattern belonging to not-so-smart strategy: randomly choosing and placing speakers, resulting in still to be scheduled speakers that do not fit anymore, and consequently rearrangements have to be made.

| | |
|---|---|
| 1 | 1 |
| 2 | 1 |
| 3 | 1 |
| 4 | 1 |
| 5 | 2 |
| 6 | 2 |
| 7 | 2 |
| 8 | 2 |
| 9 | 3 |
| 10 | 3 |
| 11 | 3 |
| 12 | 3 |
| 13 | 3 |
| 14 | 3 |

Figure 5.1 - 14 Speakers placed in a 100% ideal sequence, and without superfluous moves

| | |
|---|---|
| 1 | 3 |
| 2 | 3 |
| 3 | 1 |
| 4 | 2 |
| 5 | 3 |
| 6 | 3 |
| 7 | 2 |
| 8 | 2 |
| 9 | 1 |
| 10 | 3 |
| 11 | 2 |
| 12 | 1 |
| 13 | 3 |
| 14 | 1 |
| 15 | 3 |
| 16 | 3 |
| 17 | 1 |
| 18 | 2 |
| 19 | 3 |
| 20 | 3 |
| 21 | 1 |

Figure 5.2 - 14 Speakers placed in a 44.3% ideal sequence, and with 7 superfluous moves (21 in total)

### 5.1.5 Eye Tracking Equipment

Eye tracking was performed using the Tobii ©1750 eye tracker with a 17” LCD monitor (96 dpi) set at a resolution of 1024x768. The eye tracker sampled the position of users’ eyes at the rate of 50Hz. An integrated log of eye movement data and user events allowed us to map eye movements to various features on the screen during task performance. This application provided us with the screen coordinates for each region that we were interested in. For analysis, we used Clearview ©, which comes with the eye tracking hardware. The experiment took place at the Utrecht University Institute of Linguistics who were so kind to let us use their facilities. We will measure fixation length, which indicates how long subjects are looking at or within a certain area, and fixation frequency, referring to how often subjects look at a region (Jacob & Karn, 2003).

### 5.1.6 Procedure

The experiment took about half an hour. Subjects were welcomed, followed by a textual introduction and a small video fragment showing how the application and its controls worked. After this, subjects started to work on the 6 problem solving tasks.

## 5.2 Hypotheses

In this experiment, the focus is on more transfer tasks. In experiment 4, doing the initial phase (Conference Planner) in the Internalization interface resulted in better performance on the transfer task. However, then we could only analyze one transfer task. We are interested

to see how long (for how many tasks) this effect lasts. Besides this research question, we formulated two other hypotheses:
**H1: Internalization subjects will apply a better strategy.** Because they work in a more plan-based manner, we expect Internalization subjects to come up with more smart and efficient strategies reflecting that the constraints of speakers have been taken into account. Externalization subjects are expected to do this to a lesser extent, and therefore are expected to display less smart strategies that point more at trial and error problem solving, instigated by interface cues.
**H2: Externalization subjects will spend more time looking at screen areas where the destination feedback takes place and less at the areas where the information of speakers (constraints) are displayed.** Externalization subjects are expected to act more on a display-based manner, driven by interface cues, which take place on the right side of the screen, where we expect them to look more. Internalization subjects are thought to concentrate more on strategy construction and think more about their moves. The information needed is located on the left side of the screen, where we expect these subjects to look more.

## 5.3 Results Experiment 5

We will mention the same measures as in experiment 3 and 4, starting with the results for the initial phase, then the transfer phase and then lastly elaborate on the data from the eye tracking procedure.

### 5.3.1 Initial phase: Conference Planner

See below the results of the initial phase that consisted of three trials with Conference Planner. They were obtained with repeated measured ANOVA with the independent variables interface style and task (there were 3 tasks in each of the two phases).

Table 5.2 - Time and move based measures per interface style in the initial phase

| Average for the 5 tasks | Internalization | | Externalization | |
|---|---|---|---|---|
| | Mean | SD | Mean | SD |
| Total time needed (s) | 169.7 | 45.5 | 176.6 | 38.5 |
| Time before first move (s) | 14.7 | 5.7 | 13.2 | 5.7 |
| Time between moves (s) | 5.8 | 1.8 | 5.3 | 2.8 |
| Duration of moves (s) | 2.1 | 0.7 | 2.2 | 0.7 |
| Superfluous moves | 6.4 | 4.1 | 10.3 | 8.4 |

**Time based Measures**
*Total time needed*
As in experiment 3 and 4, the two interface styles caused no significant difference ($p > .05$) on the mean total time subjects needed per task.

*Time before first move*
Unexpectedly, the average time before the first moves per task, did not differ significantly ($p > .05$) between the conditions (unlike in experiment 3 and 4).

*Time between moves*

As in experiment 3 and 4, the time subjects took between individual moves, was significantly higher $F(1, 50) = 2.8$, $p < .05$, for the Internalization subjects ($M = 5.8$, $SD=1.8$) as compared to the Externalization subjects ($M = 5.3$, $SD = 2.8$).

*Total time needed*

As in experiment 3 and 4, the two interface styles caused no significant difference ($p > .05$) on the mean duration of moves per task.

**Superfluous moves**

As in experiment 3 and 4, in the initial phase where Conference Planner was used, the Internalization group ($M = 6.4$, $SD = 4.1$) outperformed the Externalization group ($M = 10.3$, $SD = 8.4$) in terms of less superfluous moves, $F(1,50) = 4.52$, $p < .05$, see figure 5.3.

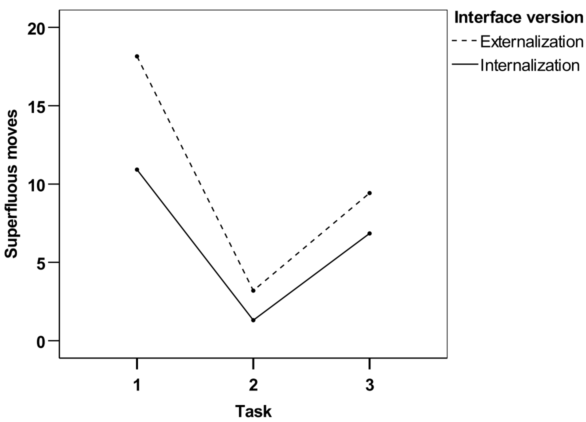



Figure 5.3 – Mean number of superfluous moves in phase 1 per task and interface version

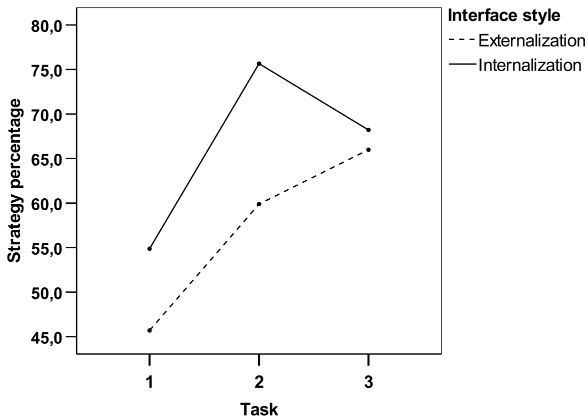



Figure 5.4 – Strategy: mean percentage with which the smartest sequence has been applied per task and interface version

In figure 5.3 (and some of the ones that follow), there is a clear effect of task visible. Task 2 was done with less superfluous moves by all the subjects. This pattern can be seen also later on in other graphs. The explanation for this, we assume, is that the to be solved constellation in task 2 was easier than the other tasks, and that this was the case for the subjects in both conditions.

**Strategy**

Regarding the strategy measure during the initial phase it appeared that interface style had a significant effect $F(1,50) = 5.11$, $p < .05$. Expressed as percentage, the Internalization group applied the smart strategy significantly more ($M = 66.3$, $SD = 14.9$) than the Externalization group (57.2, $SD = 14.00$), see Figure 5.4. There were no interaction effects. The above figure suggests an interaction effect, but there was none ($p > .05$); it showed to be interface style only that was the deciding factor causing the found differences. However, still the trend that this difference diminished over time is clearly visible; by task 3 most of the subjects probably had found out that it pays of to think moves over before making them, and it was not so hard to realize that it is easier to start with the more difficult speakers.

### 5.3.2 Transfer phase

**Performance on the transfer tasks**

The transfer phase task consists of 3 tasks. As in experiment 4, we can look at performance on the *first* transfer task, after having done three tasks with the Conference Planner in the Internalization or Externalization interface style. In addition, we are now also able to look at two next successive transfer tasks to see whether possible differences in performance are only there in the beginning, upon switching to the different task, or that the effects last longer.

**Time based measures**

There were no effects of interface version worked with before, interface version in the transfer phase itself or interaction effects concerning any of the time measures.

**Superfluous moves**

The following measures were obtained by performing repeated measures ANOVA with the independent variables interface style and task (there were 6 tasks). In the former experiment (because of the complex, not entirely orthogonal design), we could only look at one transfer task. In this experiment, there are three transfer tasks for all the subjects. There was an interaction effect that was almost significant at a conventional level of *initial* interface style (Conference planner) and interface style in the *transfer phase* (Ferry planner) on average superfluous moves in the transfer phase as a whole, $F(1,50) = 3.45$, $p < .06$. Doing the initial phase and the transfer phase in the *same* interface style (EXT→EXT, M = 3.5, SD = 1.4 and INT→INT, M = 4.3, SD = 1.2) caused no difference in performance in the transfer phase. These scores did also not significantly differ from the sequence INT→EXT (M = 5.2, SD = 6.5). However, the sequence EXT→INT scored worse than the other three combinations, M = 6.2, SD = 1.2).

**Externalization interface in initial phase**

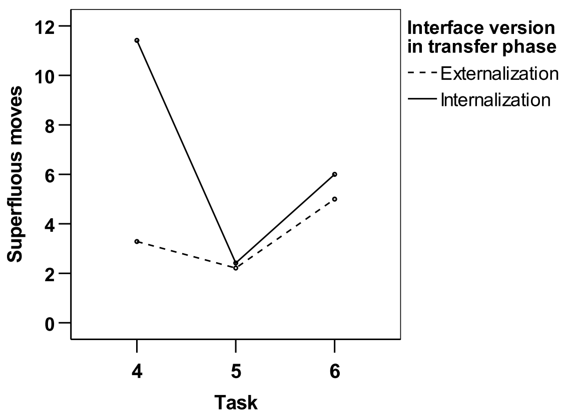



Figure 5.5 – Superfluous moves in the transfer phase per task and version of subjects that worked with Externalization before

**Internalization interface in initial phase**

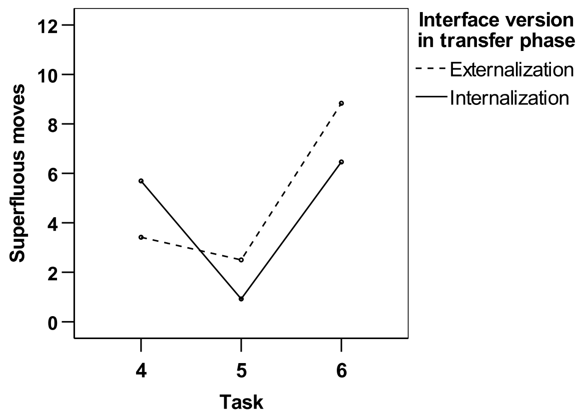



Figure 5.6 – Superfluous moves in the transfer phase per task and version of subjects that worked with Internalization before

When looking at the course of events (superfluous moves per task) in figures 5.5 and 5.6, attention is likely to be drawn at the peeking score for the sequence EXT-INT in figure 5.5. We decided to look deeper into this, and performed a separate analysis on the first transfer task. In task 4, the first transfer task, there was an interaction effect (see figure 5.7) of initial interface and transfer interface on performance, $F(1,50) = 4.75$, $p < .05$. When subjects worked with the Internalization version in the initial phase it did not matter with which interface they continued in the transfer phase. This corresponds to the results of experiment 4, also there only in the first transfer task effects were found. When working in the sequence EXT→INT, the number of superfluous moves was significantly higher than scores of the other sequences (INT→INT, INT→EXT and EXT→EXT). After task 4, this difference disappeared. There were no other significant differences.

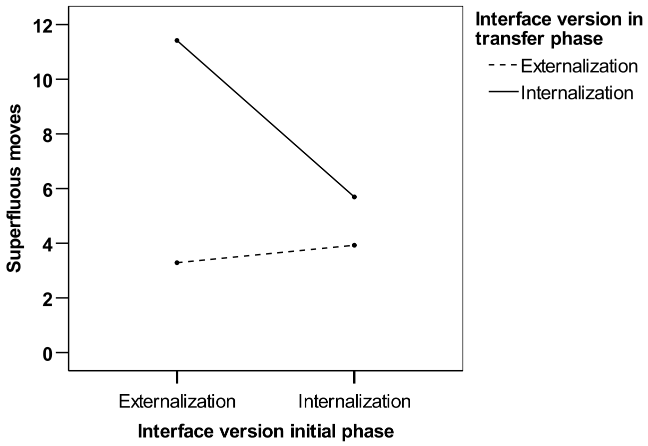



Figure 5.7 - Mean superfluous moves in first transfer task, per initial interface style and transfer interface style

*The order in which the interface styles were presented*

As in experiment 4, we can look at the influence of the order in which interfaces were switched. In this experiment, because of the design, only half (26) of the subjects did switch interface. There was again a main effect of the order of the interface switch $F(1,24) = 13.6$, $p < .001$. It showed that switching in the sequence EXT→INT resulted in more superfluous moves ($M = 10.7$, $SD = 7.5$) on the transfer task than switching in the order INT→EXT ($M = 3.42$, $SD = 6.8$), see Figure 5.8.

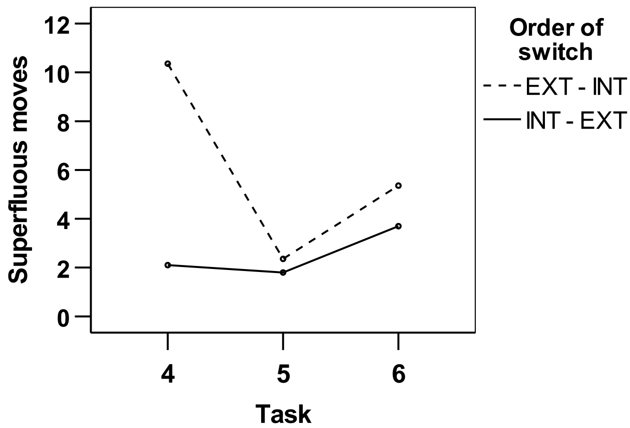



Figure 5.8 - Mean superfluous moves: Order of interface switch and when it took place

There was also an interaction effect, $F(1,24) = 4.5$, $p < .05$ of switch order and task. Post hoc analyses revealed that subjects that received the interface switch in the order EXT→INT performed worse than the ones that had INT→EXT, but only in the first transfer task (task 4), after that the difference was gone.

**Strategy**
By phase 2, there was no difference anymore between the amounts of “smart” strategy applied, interface used in the first phase or interface version in the transfer phase had no influence anymore, and there were no significant interaction effects.

### 5.3.3 Results Eye tracking Data

**Regions of interest**
In the conference Planner application, there were areas where constraints of both speakers and rooms were displayed, and there was the area where the speakers eventually had to be scheduled (see Figure 5.9). To be able to let Clearview (analysis software) provide output in connection to certain screen regions, we divided the screen into several regions as can be seen in Figure 5.10.

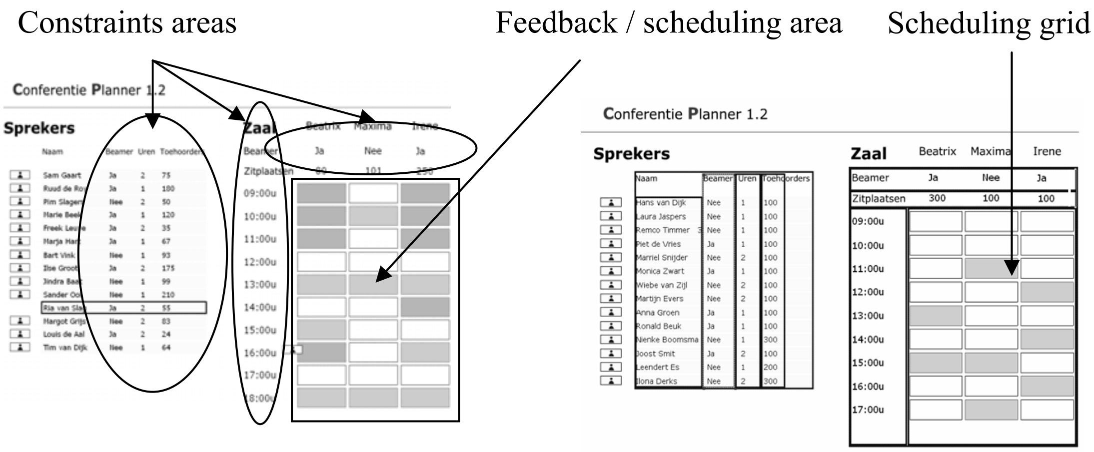



Figure 5.9 - Conference Planner and the different regions

Figure 5.10 - Defined Regions of interest in Conference Planner

Since we assume that attention is being drawn to the green destination feedback in the scheduling grid, the scheduling grid is an important area to examine with eye tracking data. We expect that Externalization subjects spend more time looking there, since they are expected to work not so much based on a plan, but more in a trial and error fashion. Trial and error would implicate that speakers are being picked up on the left side without much planning, and then again without a lot of thought placed on the right side on a slot that is “green”. The green destination feedback is at work from the moment on that a speaker has been clicked or picked up, hence we expect that Externalization subjects’ attention will be drawn to the scheduling grid more than is the case with Internalization subjects.

**Phase 1**

For task 1, 2 and 3 with the Conference Planner we compared two eye tracking measures between Internalization and Externalization subjects. The first one is the average fixation length per task and per region, which is the average *total* time that users spent looking (gaze) at an area of the screen. Secondly, the average number of fixations per task was measured, which refers to the number of fixations on a certain region. The threshold was set at 100ms, meaning that when a user fixated 100ms or more it is counted as a fixation (see also Inhoff & Radach, 1998).

From the whole range of regions of interest from the figures above, concerning one region we found significant effects. There was a significant difference in the amount of time spent looking (fixation length) at the *scheduling grid* (see Figure 5.9) of the Conference Planner application in the initial phase $F(1,50) = 4.7$, $p < .05$. On average, Externalization subjects spent more time ($M = 53.1s$, $SD = 5.2s$) looking at the grid than Internalization subjects ($M = 36.3s$, $SD = 5.6s$), see Figure 5.11. However, closer analysis showed that these differences seem to be largely caused by an almost significant interaction effect $F(1, 50) = 2.7$, $p < .07$ for interface version and the task (1, 2 or 3). Post hoc analysis revealed that Externalization subjects looked longer time at the grid than Internalization subjects in task 1 and task 2, but not anymore in task 3.

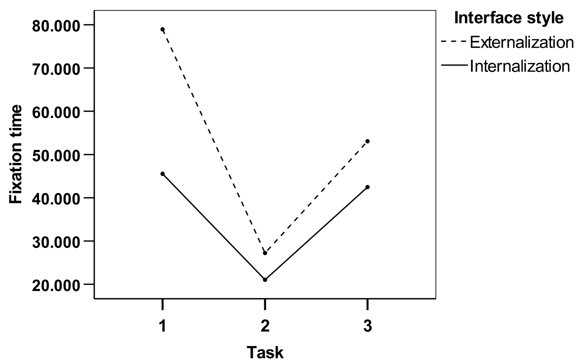



Figure 5.11 – Mean total fixation time (ms) on the scheduling grid during the initial phase

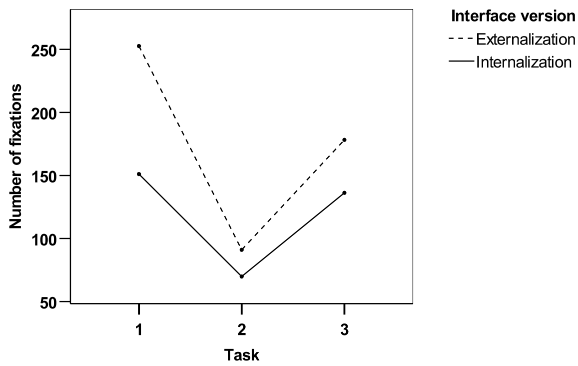



Figure 5.12 – Mean total number of fixations on the scheduling grid during the initial phase

The same pattern was found for the number of fixations (figure 5.12). There was a main effect, $F(1,50) = 6.2$, $p < .05$ of interface style on the number of fixations in the scheduling grid. Externalization subjects looked at it more often ($M = 173.9$, $SD=15.0$) than Internalization subjects ($M = 119.1$, $SD = 16.2$). Also here, the differences seem to be influenced by a tendency of an interaction effect of interface version and the task (1, 2 or 3), $F(1, 50) = 2.21$, $p < .10$. Post hoc analysis revealed that Externalization subjects looked more often at the grid than Externalization subjects in task 1 and task 2, but not anymore in task 3.

**Further measures and Phase 2**

For Phase 1 we also examined the same measures, fixation length and fixation frequency between the moment that the puzzle came onscreen, and the first move was made, but no significant differences were found there. In Phase 2, there were no differences anymore between eye movement behaviors between the different interface styles for any region.

## 5.4 Conclusions

The results of experiment 3 and 4 were largely replicated. In the initial phase, starting with a series of Conference Planner Tasks, Internalization subjects showed smarter solutions by solving the problems with greater efficiency (fewer superfluous moves). Also on the most time measures results were the same as before. In general, once more, the Externalization condition did not result in any advantages; it resulted in worse performance in terms of efficiency of the solutions.

An important reason to conduct this experiment was that we wanted to be able to see what happens later on in a transfer phase, since this was not possible in experiment 4. In this experiment, there are three transfer tasks for all the subjects, half with the same interface as before, and half with the other one. Concerning transfer, the same findings as in experiment 4 applied; receiving a new interface in the order EXT-INT (being used to guidance and then suddenly working without) resulted in worse performance than any other combination. The result was the same as in experiment 4. We now looked also at the course of events that followed, and it showed that the effect disappeared after the first transfer task. Although users in the EXT-INT sequence were confronted with a new interface that demands more mental effort, the changes in task or task environment (the interface) are mastered quite quickly. We conclude that this effect of interface style seems to last shorter than in the initial phase, which we attribute to the fact that it was in the initial phase where the most learning took place. By task 5, the second transfer task, subjects quickly recovered.

A second reason for this experiment was our desire to have a more quantifiable measure for the strategy subjects used. Having a continuous measure that reflects the potential smartness of a sequence of moves, proved to be fruitful. In our tasks speakers could be scheduled in 1, 2 or all 3 of the available rooms and speakers were thus labeled 1, 2 or 3 (indicating decreasing constraints). A smart strategy takes constraints in account, and starts with scheduling speakers with many constraints who can be placed in only one room. We analyzed to which extent subjects applied a strategy with the most-constraints-first approach (schedule speakers in a 1-2-3 sequence). The strategy measure is expressed as a percentage in which the 1-2-3 sequence in scheduling took place. In the first two tasks, right from the beginning the Internalization subjects showed smarter sequences of moves, and this of course fits well with the finding that they needed less superfluous moves. By the third task, the scores on the strategy measures converged, and in the transfer phase, earlier difference in scores had disappeared all together. This in turn coincides with the absence of differences in superfluous moves in later tasks. Subjects in the Externalization condition might have realized that it is smart to apply some sort of strategy and not do it randomly (as the Externalization interface invites subjects to behave). This insight came by task 3, and later, also during the transfer tasks, all the subjects carried on, using more or less the same consistency in placing difficult speakers first. In the context of CLT, we assumed that having to internalize information oneself, goes hand in hand with high Germane Cognitive Load. The evidence that smarter strategies were applied by the Internalization subjects, answers a part of our main research questions (section 1.5.1), and supports the first hypothesis of this chapter, which until now, except for some indications in experiment 3, was mainly based on efficiency measures. The Externalization subjects who were assumed to experience high Extraneous Cognitive Load (which hinders strategy construction) indeed showed less smart strategies in the first two tasks.

A third reason for this experiment was that we wanted our assumptions to be backed up by eye tracking data. Regarding durations of fixations and the frequency of fixations, we found that in task 1 and 2 the Externalization subjects spent significantly more time looking at the scheduling grid on the right side, but there were no significant differences for any other screen region. One then could expect a corresponding difference (in the other direction) on the left side where the speakers were listed (since overall time did not differ) but this was not the case. Perhaps Internalization subjects' fixations were more spread over the different areas. Still, our second hypothesis can be accepted to some degree, since Externalization users indeed invested more effort in looking at areas where information needed for proper strategy construction is *not* found. They invested more effort in looking where externalized information provided by the destination feedback was, and since this was not accompanied by better performance or strategies, we are strengthened in assuming that indeed more display-based behavior was instigated by the Externalization interface.

In the two first tasks, Internalization subjects, although they made fewer superfluous moves and scored higher on the strategy measure, did not spend more time looking at information on the left side than Externalization subjects. We did not expect this. We have to conclude that where subjects look in a task context like ours, and how often and for how long, does not necessarily say something about the thought process that takes place in the meanwhile. Having that said, we still assume it is safe to at least conclude that when subjects were looking at the scheduling grid on the right side (which Externalization subjects did more), they were *not* studying and pondering the constraints that were listed on the left, on the basis of which it is easiest to develop a smart strategy.

A fact is that subjects in our two conditions looked at the information on the *left* side on the screen equally often and for the same duration. A thought in spite of this, is that the mental processes that took place could still have been different (as reflected by the different strategies applied), however, our eye tracking data cannot backup this assumption. Nevertheless, we can theorize about which course of events could have taken place.

*Internalization* subjects are assumed to approach the problem and study the constellations in a plan-based manner, and make choices as to what to do. What they are thought to do more than Externalization subjects, is applying strategies in which the hardest speakers were place first, which indicates a smart strategy. Their actions are more determined; they look at relevant information, choose a speaker on base of their strategy, and place the speaker without much doubt, spending less time on the scheduling grid on the right. They have better solutions in terms of strategy, but also in terms of directness of their path (superfluous moves)

*Externalization* subjects are thought to act differently in that they partly rely on, and are guided by the externalized information which gives them the feeling they are being assisted. This could make them act more shallow, and cause them picking their speakers in a more random fashion (as indicated by strategy measures), drag them to the right side, and take more time *there* in *deciding* where to put the speaker (perhaps distracted, or confused by the green feedback?).

These are courses of events, we theorize, which could account for more time looking at the scheduling grid on the scheduling grid on the right side. This in itself makes sense: we assume Internalization subjects to think less shallowly. This means to not only think smart before you pick up a speaker, but also think *before* you place him, or even better: already knowing where to place him. Another important reason that must be at play also, is the number of superfluous moves that were made. More moves means more time, and many of the superfluous moves are rearrangements *within* the grid on the right side.

Resuming, as to our specific research questions, question 1 concerned how problem solving performance evolve over time was addressed. Several earlier findings indicating advantages for the Internalization version were replicated and fortified by measures showing better strategies as well for these subjects. Also eye tracking confirmed this partly. Regarding question 2 concerning transfer of skill to a *similar task* and question 3, also involving transfer, focused on the influence of interface style on performing a task afterwards *with another interface style*. Again, we found better performance when the Internalization interface was used first, the results of experiment 4 were replicated. However, this effect did not last longer than one task.

We can connect these findings back to CLT. The Extraneous Cognitive Load that we assume Externalization causes, can be seen as reflected by the fact that they spend more time looking around the scheduling grid, while their solutions were less efficient, and their strategies reflected less elaboration. Having to internalize information, being provoked to invest mental effort by not having guidance might allows Germane Cognitive Load to be high. The Internalization subjects were not distracted by externalized information, kept the focus more on mental processes instead of spending too much time looking at dynamic visual feedback that might give a user a feeling he is being assisted, but in the end brought no advantage at all.

# 6 Discussion

In this chapter, we discuss the findings and implications from the series of experiments elaborated on in chapter 3, 4 and 5. This research concerned interface presentations (with, or without Externalization) and the conditions under which motivation and deeper contemplation for the task can be provoked to achieve better task performance, fewer errors, and having users who continuously are on top of the task. We aimed to obtain better insight into plan-based vs. display-based problem solving and the role Externalization and Internalization play in this, and to acquire more knowledge about the mental processes involved, including their strengths and weaknesses. Several controlled experiments were conducted using two interface styles: one in which certain information is externalized onto the interface (Externalization) and another one where this is not done (Internalization). The central assumption was that Externalization of information by providing visual destination feedback, although fewer rules have to be learned and retrieved from memory, leads to shallower processing, less proactive contemplation and less metacognitive activity. An interface where users have to *internalize* certain information by themselves without guidance or assistance, might seem an unfriendly system, but was expected to result in more efficient solutions and better imprinting of task knowledge in memory. We investigated the conditions under which *externalizing* interface information influences users' performance in solving problems requiring planning. In tasks where planning is required we aim to specify which interface presentation (with or without Externalization) leads to the best performance.

We discuss our findings and their implications on the same three levels as the research questions fall into (section 1.5). The first level concerns behavior and how interface style (Internalization or Externalization) influences performance on various measures in tasks requiring planning. Secondly, there is the theoretical level, concerning implications for the theoretical concepts and frameworks to which we connect our research. The third level involves the practical relevance of our findings. An interaction framework concerning when and where to apply Externalization features (or not) in the interface will be proposed. We end the chapter with a discussion on how game research and our line of research could mutually profit from each other.

## 6.1 Behavioral level

In this section, we will answer our research questions by resuming all the results of the two series of experiments. Our main research questions regarding behavior and performance were:

- In tasks where planning and deeper contemplation is required, which interface style leads to more plan-based behavior, better strategy, and consequently better task performance?
- Besides immediate performance, which interface style causes the best knowledge about the tasks, rules and solutions afterwards (shortly after, and after a longer delay)?
- Which interface style leads to the best performance in a transfer situation (altered task or interface circumstances) and better resistance against a severe task interruption?

To be able to address the main questions, we answer the following specific questions first (section 1.5):

1. In repeated problem solving tasks, how does problem solving performance evolve over time?
2. In the case of transfer of skill to another task, does learning and doing a task in one interface style have an influence on performance on a similar (but different) task?
3. With transfer of skill to another interface style, does learning and doing tasks in one interface style have an influence performing that same task afterwards with the other interface style?
4. What is the influence of instruction to the user before the tasks (instruction to plan carefully, vs. telling users that mistakes are not a problem) on performance with the two interface versions?
5. Do individual differences in cognitive style play a role in this?
6. Delay: do the two interface styles have different effects on performance after a large delay?
7. Do interruptions of the problem solving process have different effects in different interface styles?

In two series of controlled experiments (5 in total) with different applications and tasks, we manipulated destination feedback in the interface (the Externalization interface had this, the Internalization version lacked it). In the first series with the Ball & Boxes application, the information that was externalized informed users of the applicability of using an operator to make a move. This information was externalized *continuously* in a (rather abstract) puzzle task. In the second series with the Conference Planner and Ferry Planner applications, which were more towards everyday realism in software applications, the legal options to move speakers were highlighted in green and thus externalized, but only *when a user clicked* on them (as in the guidelines by Sun and Apple).

For each of the two series of experiments we will summarize and discuss the most important findings and connect them back to the hypotheses that belonged to each experiment. We report on various measures, but the most important measures for us are those reflecting planning and strategy: the solution path (superfluous moves) and the strategy itself. Also other measures were collected, among which timestamps, user opinions, and task knowledge afterwards. These measures were obtained by computer-recorded log data (various mouse-related events), eye tracking data, and questionnaires. In the next section, we will discuss the results per experimental series.

### 6.1.1 Findings from the Balls & Boxes experiments

Most important findings:

- Generally, the Internalization subjects solved more puzzles (nearly significant main effect)
- Only in the beginning, Externalization subjects needed slightly less time to solve the puzzles
- Overall time needed for the task was practically the same in the different interface versions
- After a break, Internalization subjects perform less superfluous moves (unnecessary moves)

- Generally, the Internalization version yielded better knowledge afterwards
- After a large delay, Internalization subjects had better memory for the same task
- After a large delay, Internalization subjects performed better on a transfer task
- When a planning instruction was given, only the Internalization subjects followed it

Our hypotheses assume that an Internalization interface (which may seem less friendly than could be) can be beneficial. In general, there was the tendency that Internalization subjects solved more puzzles than the Externalization subjects did. The first hypothesis from experiment 1 session A (section 3.3.5) was that Externalization subjects *initially* perform better than Internalization subjects. This was true only for the time subjects needed; on other measures there were no differences. We conclude that more time points at Internalization subjects investing more effort in actively trying to come up with a solution (which they did, they solved more puzzles correctly). This plan-based behavior is in contrast with a more display-based behavior, which we presume the Externalization subjects to apply. The second hypothesis from experiment 1 session A stated that Internalization causes better task performance later on, especially after a pause. After the pause, Internalization subjects indeed made less superfluous moves than Externalization subjects did, and they also reached less dead-end-states (reflecting better insight in the problem space). We ponder (knowing that in the beginning there were not many differences) that Internalization subjects who did more mental processing indeed imprinted strategies better. Consequently, after a pause, they have the solutions more ready available. The third hypothesis in experiment 1 was that Internalization subjects would have better knowledge afterwards. There was a tendency confirming this regarding procedural knowledge, and for declarative knowledge about the most important rules there was a main effect. We conclude that doing more mental effort oneself was the reason for better knowledge of rules and solutions.

In experiment 1 session B (section 3.4.1) that took place after 8 months, the same subjects were re-invited and solved the B&B puzzles again. Our first hypothesis stating that Internalization subjects will have better memory of the rules and solutions was accepted. It took the Internalization subjects only half as long as the Externalization subjects to solve the first B&B puzzle. Finding differences after a long time interval confirms our expectation that the Internalization version caused users to contemplate more, and consequently have better recollection of the procedures and rules. A second hypothesis was that Internalization subjects perform better on a (different) transfer task. The Internalization subjects solved it the first time faster and moreover, they managed to solve the puzzle more often in the same time span. The transfer task being done better by Internalization subjects strengthens us in the belief that demanding more processing effort from the user causes better-imprinted knowledge.

In experiment 2 (section 3.5) we analyzed whether giving a planning instruction (low or high) to subjects influences performance, and especially whether the instruction has different effects in the two interface styles. We hypothesized that the negative effects of the Externalization version are lessened by explicitly stimulating subjects to think deeply, but that when subjects are instructed to approach the problem in a shallow manner, performance might be even worse than without instruction. We also hypothesized that the Internalization subjects might be influenced to a lesser degree, because their intrinsic motivation is thought to be higher already. It showed that the Externalization subjects were deaf to this instruction, whereas the Internalization subjects followed it. Only in the Internalization version, a high planning instruction resulted in solutions that were more efficient. We conclude that in the Externalization version, the feeling that assistance is provided, combined with the mental effort that has to be devoted to the perception of the destination feedback, absorbs part of the focused attention that is needed to follow the instruction and paying attention to it.

### 6.1.2 Findings from the Conference Planner and Ferry Planner experiments

Most important findings:

- Again, there were no *overall* time differences
- Internalization subjects take more time before they start solving the problems, and take more time between the individual moves (reflecting planning and contemplation)
- Internalization subjects perform less superfluous (unnecessary) moves (better efficiency)
- Preexisting cognitive style, measured by the Need for Cognition scale, had no influence on behavior
- The Internalization subjects are better at task resumption after a severe interruption
- Switching interface in the sequence INT→EXT results in better performance than EXT→INT
- Switching from initial task to transfer task in the sequence EXT→INT causes worse performance than other combinations
- Internalization subjects applied more efficient strategies as measured by a formal strategy measure
- Although earlier findings were replicated, the eye tracking data itself (apart from the log files we used all along) did not show much difference in gaze patterns between the two conditions
- The only differences were that in task 1 and 2 the Externalization subjects spent significantly more time looking at the scheduling grid on the right side, both in frequency, and duration

Again, we generally found that not on any level the Externalization subjects were at advantage, and again there were no overall time differences. The first hypothesis of experiment 3 (section 4.4.4) stating that the Internalization interface leads to a more plan-based strategy and better performance than Externalization, was accepted. Internalization subjects took more time before starting to work on the problems and took more time between moves, both indicating contemplation and planfulness. Internalization subjects made less superfluous moves (more direct solutions, greater efficiency). There was also the tendency that Internalization subjects worked according a better strategy, they used the "most constraints first" strategy more often than Externalization subjects did. Internalization subjects also tended to score higher on questions concerning declarative knowledge. Lastly, we asked whether it sometimes happened that they did not know how to proceed with the arrangement of the speakers. The Internalization subjects reported to have experienced this to a lesser degree. Resuming, the findings concerning time indicated effort to contemplate and think the situation over, which is confirmed by Internalization subjects having better solutions. They also applied a most-constraints-first approach more, which is the safest way to a smart and economical solution. In addition, Internalization subjects' own opinion of being stuck less in the process fits nicely with it. The second hypothesis of experiment 3 assumed that subjects who have a cognitive style reflecting a tendency to engage in, and enjoy effortful cognitive tasks (as measured by the Need for Cognition Scale) apply a more plan-based strategy than subjects scoring low. The hypothesis had to be rejected: it was interface style only that accounted for the found differences. The Need For Cognition had no influence. We conclude that the mechanism of providing guidance discourages active contemplation, is so strong that it overrules individual characteristics of users.

In experiment 4, we looked at interruption and transfer. The first hypothesis (section 4.5.7) stating that users of the Externalization version are affected by the interruption more because they work in a shallower manner was accepted. After a severe interruption, Internalization subjects continued as if nothing happened, whereas the performance of the Externalization subjects deteriorated badly. We conclude from this that Internalization subjects have worked in a more plan-based manner and have put more effort in forming strategies and solutions. This causes better resistance to a severe interruption; they can continue based on internal plans and knowledge that are better imprinted and retrievable. The second hypothesis concerned only the situations where task transfer took place (be it in the same, or in the different interface). We expected that in general, subjects who did the initial task with the Internalization version in the initial phase would perform better on a transfer task. This hypothesis was only partly accepted. It showed that doing the initial task and transfer task in the sequence INT→INT, INT →EXT, and EXT→EXT performed equally well. However, doing the initial task in EXT and the transfer task in INT resulted in worse performance than the other three sequences. This means that when *task* transfer is likely to occur, this is safer within the same interface style, or from INT→ EXT. The third hypothesis concerned those situations in which switching interfaces (INT→EXT or EXT→INT) took place (be it in the same task or a transfer task). We expected that performance after switching interfaces in the sequence INT→EXT will result in better performance than switching in the sequence EXT→INT. Switching to an interface in the sequence INT→EXT indeed resulted in fewer superfluous moves than in the sequence EXT→INT, regardless of whether it was the same task or in a transfer task, so also this hypothesis was accepted. This indicates that when having learned the task in a situation where one has to invest more effort and think for oneself, results in better knowledge of the rules and solutions. Suddenly switching to an interface where guidance is provided does not affect users, they have their routine already, in a flexible way. When working in the Externalization version where assistance is provided, part of the processing is done by the system (a situation where primary and secondary task are intertwined as in section 1.4). The user invests less effort, and consequently has less command over the procedures and solutions. When suddenly provided with an interface that lacks this information while the user is still “counting” on it, leads to decreased performance. When interface transfer is likely to occur, it is unwise to let this happen in the order EXT→INT.

In the last experiment (see chapter 5), we used the same material as in experiment 4 and the previous results were replicated. In addition, we had a closer look at the strategy applied by subjects, a closer look at transfer and we applied eye tracking to examine whether we could find differences in where subjects looked and at which frequency and duration. The way we analyzed the strategy applied here was more formally, and it indeed showed that what we presumed was the case: better strategies by Internalization subjects. They applied a most-constraints-first approach more, which is the safest way to a smart and economical solution. The order in which they scheduled speakers reflected that they started with speakers with many constraints first; they did this more than Externalization subjects. We conclude that a more plan-based approach is indeed taking place, leading to smarter, more elaborated strategies. Now also more formal strategy data contribute to evidence regarding our notions. The way we looked at transfer was different in that there were more transfer tasks, so we were be able to study a sequence of tasks to see whether the presumed transfer effects prolonged. Again, as in experiment 4, doing the first transfer task with the *Internalization* interface after having worked with the Externalization interface before resulted in more superfluous moves. However, this effect disappeared rapidly. With a task such as Conference Planner, it showed that the effects are especially found at the beginning of a series of trials. Also in experiment 4 where the interruption took place, we saw that performance converged after a few tasks (until the interruption). In this experiment, regarding transfer, the findings of experiment 4 were replicated, but no further longer lasting effects regarding transfer were found. The effect disappeared after the second transfer task, from which we conclude that the disrupting effect of changing interface in the sequence EXT→INT is rather temporary, users in that situation quickly realize that another way of working is required, and they adopt a more plan-based strategy. The results of applying eye tracking did not bring as much as we hoped for. Aside from the above findings obtained by using log data, there were only few significant differences in the amount of gaze time spent looking at the *scheduling grid* of the screen in Conference Planner, and this occurred only in the beginning. This was a screen area where little relevant information could be found. Externalization subjects looked there for longer time than Internalization subjects did, and the same was true for the number of fixations. The important information that could help users construct a strategy was to be found mostly in other areas. There were no significant differences for other screen regions. We conclude that Internalization subjects performing better is caused by different mental processes (instigated by our interfaces) but that eye tracking data in itself are here not helpful in obtaining *extra* proof. The different mental processes and behavior are not to a high degree accompanied by very different eye gaze patterns.

## 6.2 Theoretical level

In this section we elaborate on our findings on a theoretical level. We also connect our findings back to research findings of others. At the end of section 1.4, we sketched some possible positive and negative effects of Externalization by destination feedback (using a chess example). The first possible positive effect was *computational offloading*, because several mental processes can be skipped. We cannot backup by measures whether computational offloading took place, but we assume it did, since we can assume that these subjects have done less mental effort, which was not very plan-based. The second suggested positive effect was that destination feedback might suggest moves the user did not think of. We cannot know for sure where users thought of, but we do know that destination feedback did not lead to better solutions and strategies, which indicates little plan-based behavior. The third possible effect suggested that because of repeated exposure to externalized information subjects could *learn the rules* of the problem. For that, we have strong indications that it was the other way around; the way Externalization subjects approached the problem points at little plan-based problem solving backed up by a solid knowledge.

The first possible negative effect we mentioned was that destination feedback could provoke laziness and shallow behavior because of the feeling that “the work is being partly carried out for the user”. We are convinced that this was indeed the case, backed up by results concerning solution paths and thinking times. The interface guided decisions, indicating display-based behavior. The second possible negative effect was that destination feedback did not aid strategy construction. From our strategy measures that indicate how subjects performed actions taking constraints in account, we had this confirmed. Externalization subjects applied less smart strategies, which suggests that the interface information guided decisions, again pointing at more display-based, and little plan-based behavior. The third possible negative effect we thought to be distraction, since the visual feedback still has to be processed. The user cannot ignore to see it, and attention will be drawn to the green squares. The feedback suggests moves by showing what is possible, and subjects are seduced to perform an action based on that, pointing at display-based behavior. That behavior was not very plan-based, is also backed up by the findings that they were deaf to instruction, and that preexisting attitudes to problem solving as measured by NFC had no influence; it was our manipulation only that caused the performance differences.

**Theories combined**

We account for our findings with a combination of explanations, namely notions from Cognitive Load Theory (effort, motivation, distraction, and attention), theories involving cost-benefit trade-off, while still accepting notions from Information Processing Psychology for the actual computational processing part as well (see Figure 6.1). We use CLT as a way to look at how certain behavior can be provoked; we do not see it as a full-fledged theory of cognition. CLT in our context shifts away from IPP approaches, but our take on CLT also slightly shifts away from CLT itself. In the whole context, when a user sits down in a given environmental circumstance, is encountered with an interface and starts with his/her task, CLT is a good starting point (therewith not denying other theories in addition) to look at behavior. Critical to task execution is the motivation users have, and the interface is the crucial part between the user and the system. Regarding the interface as the secondary task that should be optimized to facilitate the primary task, paves the way for the thought that this secondary task should be “cognitively” informed (Van Nimwegen, Van Oostendorp & Tabachneck-Schijf, 2006) in the sense that it circumvents certain human tendencies in computer behavior.

### 6.2.1 Cost-benefit trade-off

The presumed advantage for Externalization in the beginning as we hypothesized in section 3.3.5 did not show; it was the other way around. In the Conference Planner tasks, especially in the beginning, Internalization subjects were at advantage. It was only after several trials that performance started to converge between the two conditions. A cost-benefit trade-off seems to play its part in both conditions, but at a different point in time. The trade-off that takes place between cost and benefit can involve the following: a cost would be to invest effort in meaningful deeper mental processes (Germane Cognitive Load); a benefit then would be the fact that solving the problem will be plan-based and more efficient (with fewer unnecessary actions). When using an Internalization version (no externalized information) the task is seemingly harder, and subjects can be expected to think deeper in the context of a trade-off, as in the cost-benefit analysis as stated by O’Hara and Payne (1998, 1999). This showed to be the case in our Internalization versions where Externalization (by destination feedback) was left out. Users see the problem, and then quickly realize that aimlessly clicking is not going to help and interact in a plan-based manner. They make a trade-off, and realize that investing some mindful effort (cost) will pay off in the end (benefit). In the Internalization versions in our experiments, users might be aware of this right from the beginning, since they never saw any assistance or guidance. They apply a plan-based problem solving approach instead (Vera & Simon, 1993), in which they actively search for an adequate strategy. In the Externalization condition, subjects have the impression that the guidance from the interface will help them, especially in the beginning. Their behavior can be characterized as display-based problem solving behavior (Larkin, 1989). Only after several trials these subjects realize that they are not working very efficiently (they are stuck quite often) and that it might be worthwhile to invest some effort (cost) in coming up with a better strategy (benefit). The above idea would imply that when *immediate* error free performance and behavior as efficient as possible is desired, one might be cautious with designing a system in which the primary and secondary tasks are intertwined in such a way that users feel they can take it easy and partly rely on the system.

### 6.2.2 Cognitive Load Theory

While embracing the concepts of CLT, we also challenged some assumptions of it, namely the basic assumption the Cognitive Load should be always kept low (see also Schnotz & Kurschner, 2007). Is cognitive load good or bad for learning? It turns out that sometimes it is good (germane) and sometimes it is not (extraneous). How to tell which from which is one source of problems with CLT. For our research issues, CLT's concepts are valuable, and in our opinion, the scenarios we postulated in section 2.6, concerning the Internalization or Externalization in context of CLT components, are adequate. We proposed that:

1. Externalization (in our case by destination feedback)
   a. takes resources away from the task itself, thereby increasing Extraneous Cognitive Load
   b. causes Germane Cognitive Load to decrease because of
      - less bandwidth for Working Memory (Extraneous Cognitive Load is high)
      - the suggestion of assistance makes users less willing to invest effort and to make inferences (little plan-based behavior)

2. Internalization (the absence of destination feedback)
   a. causes low Extraneous Cognitive Load, since no resources are (unnecessarily) taken away from the task and users are not distracted by processing interface events such as feedback (little display-based behavior because users do not see move-options displayed in the interface which could deviate them from a direct plan of their own)
   b. provokes plan-based behavior. Germane Cognitive Load is high, because users are triggered to look for underlying rules. Users are more willing too invest mindful effort in making inferences, and have more motivation to look for what underlies the tasks

Regarding the issue of the separation or intertwining of the primary ands secondary task, having the interface (the secondary task, see section 1.4) as straightforward as possible in the sense that it stays close to the essence of the task, as in the chess example, might be a smart thing to do. The availability of a wide range of options regarding interactivity does not mean that we necessarily do have to apply all the possibilities that exist. Regarding the three components of CLT, so far we incorporated mindful, effortful processing and metacognitive activity in Germane Cognitive Load. In this context, it is interesting to mention some interesting approach to improve CLT theory by Valcke (2002). He introduced *metacognitive load* as a fourth type of cognitive load and took activities that are beyond cognition into account (just as we attempted). His starting point was that monitoring plays a significant role and influences different parts of the learning process such as the selection and organization of sensory information to working memory. Since monitoring activities refer effort to process and comprehend information, Valcke incorporated metacognitive load in the category of Germane Cognitive Load.

For HCI in general, we think that CLT is a fruitful theory that can capture exactly what HCI is about: humans and their tendencies and habits to approach situations, as are the above ideas of cost-benefit trade-off. They can help us decide what is wise, and what is unwise to do. User effort and motivation, an often-deciding factor in the success of activities we perform, are sufficiently acknowledged in CLT. As said, we think IPP is not suitable to capture the complete sequence of user action and mental processes such as they are in many man-machine situations today. Then, in a computer-based task, once the right user attitude and motivation for the task has been reached, notions from IPP concerning computational issues and inferences that are made can still apply.

### 6.2.3 Germane Cognitive Load: Motivation and effort

The reason for the better strategies, better efficiency, and better knowledge in the Internalization condition is assumed to stem from the fact that these users did more proactive contemplation and generally from the start have invested more, or a different kind of effort in their (more plan-based) approach to solve the tasks. Finding out something for yourself leads to more involvement in the task. This leads to better performance during the task, better knowledge after the tasks, better transfer of skill and less vulnerability to a severe interruption. We thus conceive having better attention for the task, and investing more effort in it as the main reasons for the better performance in the Internalization version. The issue then becomes how to seduce users into this kind of behavior, for which we need methods to prime users to actively draw inferences themselves. In CLT terms, this is what would correspond with promoting Germane Cognitive Load, mental effort that is associated with motivation, meaningful thoughts that contribute to schema acquisition and the formation of proper strategies. GCL can only be high when Extraneous Cognitive Load is low (see next section). Thus, ECL being as low as possible is a necessary requirement, although only reducing ECL does not cause users to spontaneously engage in activities associated with Germane Cognitive Load (Gerjets, Scheiter, & Catrambone, 2004). Therefore, alongside minimizing ECL, we should increase Germane Cognitive Load by itself. The earlier mentioned cost-benefit trade-off is thought to partly do this: make users contemplate about what they do. However, there are several ways to increase Germane Cognitive Load. The thoughts on leaving out information might act as a trigger to people, for elaborations on this issue see section 6.4.1.

### 6.2.4 Extraneous Cognitive Load: Distraction and attention

We are convinced that in principle, users in general tend to incline on assistance or guidance of an interface when they have the feeling that this is justified. When destination feedback is provided, interactions are more interface-driven and behavior is more display-based. We saw that the Externalization as we implemented it (destination feedback increased the interactivity of especially Conference Planner) did not result in anything positive. Besides the issue of giving users the feeling that they are being assisted (simply follow arrows or place a speaker in a green spot) the Externalization of information might also possess other characteristics.

The information that was given to users about the legality of moves might have distracted them from their own thoughts and seduces them to follow what is shown on the screen. In the case of Ball & Boxes, it meant to look for highlighted arrows, in Conference Planner it involved clicking on elements and see which fields turn green. This does not aid the construction and application of adequate strategies. Moreover, even *when* a person is not seduced to do this and intends make his own plan instead, the signals (destination feedback) still takes attention, there is no way in which users, at least subconsciously can "*not see it*". Extraneous Cognitive Load (ECL) originates from the design and layout of material and refers to processes that do *not* contribute to learning (irrelevant load). It is plausible to consider providing users with externalized information (by implementing destination feedback) as ECL. Regarding distraction, the finding that a given instruction to plan carefully was not followed by the Externalization subjects, suggests that users in the Externalization version are too absorbed by the task and the interactivity the interface offers (more than in the Internalization version).

In section 6.2.2 we mentioned that it could be desirable to have a straightforward situation so that the secondary task (the interface) does not interfere with the essence of the primary task. By taking the decision to *not* implement interactivity options such as destination feedback, one in essence minimizes the information flow from the system. This makes interaction as a whole perhaps not as friendly as one is used to, but might safeguard the separation between primary and secondary task (see section 1.4) so that the right resources (motivation and effort) are directed at the primary task.

### 6.2.5 Information Processing Psychology

We suggested that IPP does not suffice for HCI issues such as the ones we studied. The reason was that we saw it as too mechanistic and rational. Task-execution in tasks as the ones we study is influenced by environmental factors (Suchman, 1987), the tool to do the task (the interface and its behavior) and habits and expectations computer users have developed. All this has consequences for motivational and metacognitive factors that play a role during interactions, and this issue deserves more attention.

Nevertheless, one could consider the mentioned processes such as cost-benefit trade-off mechanisms and CL-related mechanisms (instigated by varying Internalization and Externalization) as *influencing* IP-related processes. The thoughts on Germane Cognitive Load are associated with effort, motivation and processes directed at seducing a user to undertake mindful cognitive processes. Extraneous Cognitive Load is concerned with allowing GCL to be high and avoiding unnecessary processing and distractions. In addition, the cost-benefit explanation that we believe plays a part, is associated with the amount, and quality of processes concerning thought, and consequently decisions to invest effort in mental activity. Both components can be seen as facilitators (or inhibitors) that determine the mental processes that follow, for which IPP remains a viable theory to explain which computational processes can take place (see Figure 6.1). In other words, these two components determine the extent to which, and quality of IP processes are involved in solving the problem.

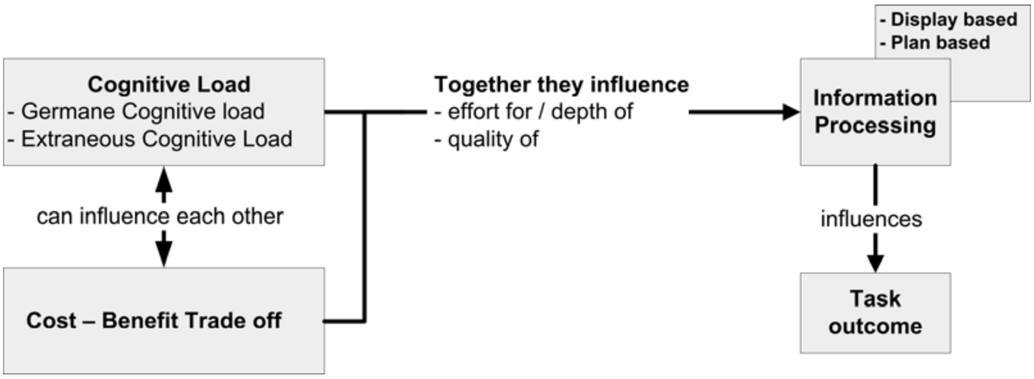



Figure 6.1 – Cognitive Load, Cost-benefit trade-off and Information Processing

The motivational processes that we explained with CLT and the cost-benefit trade-off framework both have its own combined influence on the attitude with which a user approaches his task. The final problem solving process itself cannot so adequately be described in CLT terms. Concepts of CL and the cost-benefit framework can be seen as influencing certain processes that in turn that are accommodated by IPP (see Figure 6.1). IPP is well suited to accommodate what our measures actually were about, such as path measures through a problem space and the concept of a problem space itself. Reaching dead-end-states in a problem space, for example, can be very well described in terms of IPP and the concepts of display-based and plan-based problem solving (Figure 6.1). In the Externalization interface, the interface can at some point show that performing the action to reach a certain dead-end-space is legal. A user then deciding to perform that action, is a typical example of display-based behavior, in that the display drives interactions and decisions. If the user had had an elaborated plan guiding interaction, he would never have done this, since he would know that it would get him stuck. CLT notions can be useful to account for (and hopefully predict) *why* the quality of the thinking process is different between different interface styles. Germane Cognitive Load could be seen as determining the extent of planning that precedes actions (plan-based vs. display-based), and consequently the way someone invests in rule generation or making inferences which in turn influences reaching dead-end-states. The distracting, attention-taking properties of Externalization that we associated with Extraneous Cognitive Load (seen as bad) can for example influence the clarity of operators, or the perceived simplicity of the problem space. Germane Cognitive Load, which is connected with the motivation and effort for a task, can be seen as influencing choices whether in the end trial and error, hill-climbing strategies or means-end-analysis, just to name a few concepts from IPP, will be used. In the example of the Conference Planner application, the decision of Internalization subjects to take more constraints into account during the process (visible from strategy measures) can be explained by CLT notions, but the computational process itself, is where IPP comes in.

### 6.2.6 Our findings in context of other research

Lastly, we want to elaborate on possible reasons why our results imply conclusions different from the ones of Zhang and Norman, or Kotovsky et al. Firstly, our tasks, especially the constraint-scheduling applications Conference Planner and Ferry Planner were not so much of a classical puzzle-solving nature. The interactive properties are more or less like in standard software, as was the computer hardware and the look and feel of the applications. Perhaps it was clear to subjects that they were in a "normal" computer task environment, and this might very well have seduced users to act in the way they are used to (including expectations and learned habits) more than in experiments with pure puzzle-tasks. The motivational processes that took place in our users, such as the ones influenced by a feeling one is assisted, and the fact that the Externalization happened dynamically (depending on the state in the problem space) might have contributed to the behavior, and consequently to the outcomes to a high degree. In Zhang and Norman, or Kotovsky et al.'s experiments, the tasks were pure puzzles and the interfaces were nothing like they were used to. Subjects might have reacted differently to externalized information there than in for example the Conference Planner, in that there the externalized information did *not* provoke more shallow behavior, but caused subjects to indeed actively make sense of externalized information and integrate it in a beneficial way with their internal mental process.

Furthermore, the Externalization by destination feedback as we implemented it, which is commonly used - and as far as we know has not been researched as such -, was simply different from the type of Externalizations Zhang and Norman (1994) used, where groups of objects as a whole convey static information, or not. In their experiment, the puzzle existed in three versions in which all the objects (discs, cups or oranges) constantly carried different amounts of externalized information in them. Our type of assistance goes one step further than the decision to let all the objects in the different versions of Towers of Hanoi have the same appearance. In our experiments, problem solving as a whole was more dynamical, the problem spaces larger and more branched (see Figure 3.3) and the solutions less linear.

In general, Balls & Boxes was of a more abstract nature, and there our results were most interesting after a distraction task during a pause. Zhang and Norman have not done this, although it is very interesting, because it uncovers issues related to memory for the task. As said, the issue with the Conference Planner and Ferry Planner experiments were of quite a different character both as a task, and interface-wise, and there different behavior and performance was especially prevalent in the beginning.

This brings us to a thought that will be elaborated on in section 7.4, namely that there is a need to specify more in which specific practical setting research findings such as ones mentioned (and ours as well) apply. Regarding the findings of Kotovsky et al., at first glance, their finding that their lo-info version scored better than a no-info version seems to contradict our findings. However, one has to bear in mind that the puzzle (the Chinese ring puzzle) is regarded as extremely difficult. Also in the more abstract isomorphs they constructed, the task was still quite hard. The Chinese ring puzzle lacks any common sense, and it might be simply necessary to externalize *some* information, for successful problem solving to take place in the first place. The no-info version of Kotovsky et al. really had no information, no context, no storyline, or information whatsoever. Subjects saw an abstract version of the classical Chinese ring puzzle, translated to a situation with balls and boxes (not to be confused with our application that is also called Balls & Boxes) and that was all. In our Balls & Boxes application, subjects could consult the rules, and although the solution still was unknown to them, the setting in which it takes place (such as the idea that objects are transported with some conveyor to another side) was altogether less abstract and difficult than the isomorphs of the Chinese ring puzzle. In the Conference Planner application, the rules were even better visible; it was easy to see the constraints of the rooms etc. since they were displayed. Thus, in both our applications, the rules were not kept secret. However, how to integrate this information in a way that makes sense, and mentally develop a strategy *based on these rules* under influence of interface features is what we studied. This brings us to the issue that Externalization and Internalization as we use the terms are positioned on a continuum. Internalization as we used it, would certainly not have the value “0” if the scale would be from “0” to “10”, since in our experiments there was always some information externalized. In other words, what was no-info in Kotovsky et al.’s puzzle would probably really score “0” on the scale, and it makes sense that having *some* information as in the lo-info condition causes better results. Having that said, one can approach their and our findings from a different angle, and theorize that our Externalization condition was probably closer to their all-info condition than to their lo-info condition, and that looking at it that way, our findings can be seen as not so far apart.

What is probably most important to realize, is that what we implemented and varied in our experiments, was displaying information through the operators themselves, which is different from the manipulations in both Zhang and Kotovsky et al.’s experiments. What we in essence investigated was the effect of having operators that convey information (interface controls, the secondary task, see section 1.4) that as such become intertwined with the primary task. This is quite different from the Zhang and Kotovsky et al.’s research, and focuses more on representation issues in an applied HCI context regarding mental processes and human behavior and tendencies.

## 6.3 Practical relevance: an interaction framework

The importance of this research lies in the fact that we investigated a specific kind of Externalization, the type that is referred to as destination feedback (unlike in other research). The interface controls themselves dynamically convey (externalize) relevant task information, which informs users of specific possibilities with specific objects at a certain point in the problem space.

Assisting a user by providing externalized information that might act as guidance, had negative effects. The assisting properties might cause users to rely on interface information, instead of doing effort and contemplation by themselves. It caused worse immediate performance, knowledge acquisition and transfer of skill were worse, strategies and solutions less smart and vulnerability for an interruption higher. These findings, which go against some common sense notions, can be interesting and challenging for the HCI field, since Externalization by destination feedback is highly common in systems design.

Our aim was to specify alterations to the general guidelines as provided by several large software parties. It is possible to incorporate our findings and explanations in an interaction framework (see Figure 6.2). The framework may prove valuable in tasks where processes take place in situations *outside* the category of tasks where plain *common* *usability* is high on the list. Examples of these are ATM teller machines, informational websites meant for incidental use, electronic e-applications for the tax-office that anyone should be able to fill out without much effort, or off the shelf consumer applications for e.g. home-video-editing, etc. Our framework is applicable in more serious task-environments, such as educational software where explicit learning *itself* is the aim, situations where complex task processes are mediated via a computer interface, or in situations that involve high responsibility and potential risk or cost (aviation, industrial processes, etc.).

In Figure 6.2, the framework is laid out. One should start on the top by asking the question whether the task is of such a nature that common usability is at stake, e.g. when buying a train ticket from a machine. There are several questions to be answered once one knows the task, and from there we attempt to supply a direction as to whether the implementation of Externalization is desirable. If this is the case, and if simple, straightforward (trans) actions are what is required, there is no problem implementing Externalization to guide and assist users the way it is common and users are used to. However, if this is not the case, one can check in the gray box whether any, or a combinations of situations mentioned in the white boxes applies. There are 6 broad categories of questions in Figure 6.2 (not in a specific order), which refer to the following:

1. *Interruptions likely*. A situation where interruptions are likely to occur, and that when fast, error free task resumption is at stake
2. *Aim is learning task / educational objective*. Situations where users do not have certain knowledge yet, and in which learning a task which has a certain order of processes underlying it, or when learning in the sense of knowledge acquisition is the educational objective
3. *Transfer required*. Situations where it is foreseen that knowledge or skill of one domain must be applied to another domain, when a user (or operator) has to switch between different kinds of interface styles, in a multifaceted task, or a combination of these
4. *Operations come with a cost*. Situations where actions are costly, and where unnecessary actions should be avoided.
5. *Continuous attention/mental involvement required*. When the objective is a concentrated and focused user (or learner) and distractions have a high chance of negatively disturbing the process, or in a situation where instruction is part of the process and one wants them to be paid attention to.
6. *Deep understanding of underlying domain required*. A situation where deep understanding of underlying domain is required and present, where constant full command of processes is required. When this is the case, it is important to not discourage proactive behavior based on this solid knowledge, but to give way to processes that are thought over and constructed by a user him self.

If any of the above, or a combination of any were the case, the advice would be to take care and at least think over whether Externalization (for example as we did and is common) is desired, since according to our results, it does not bring advantage in any case.

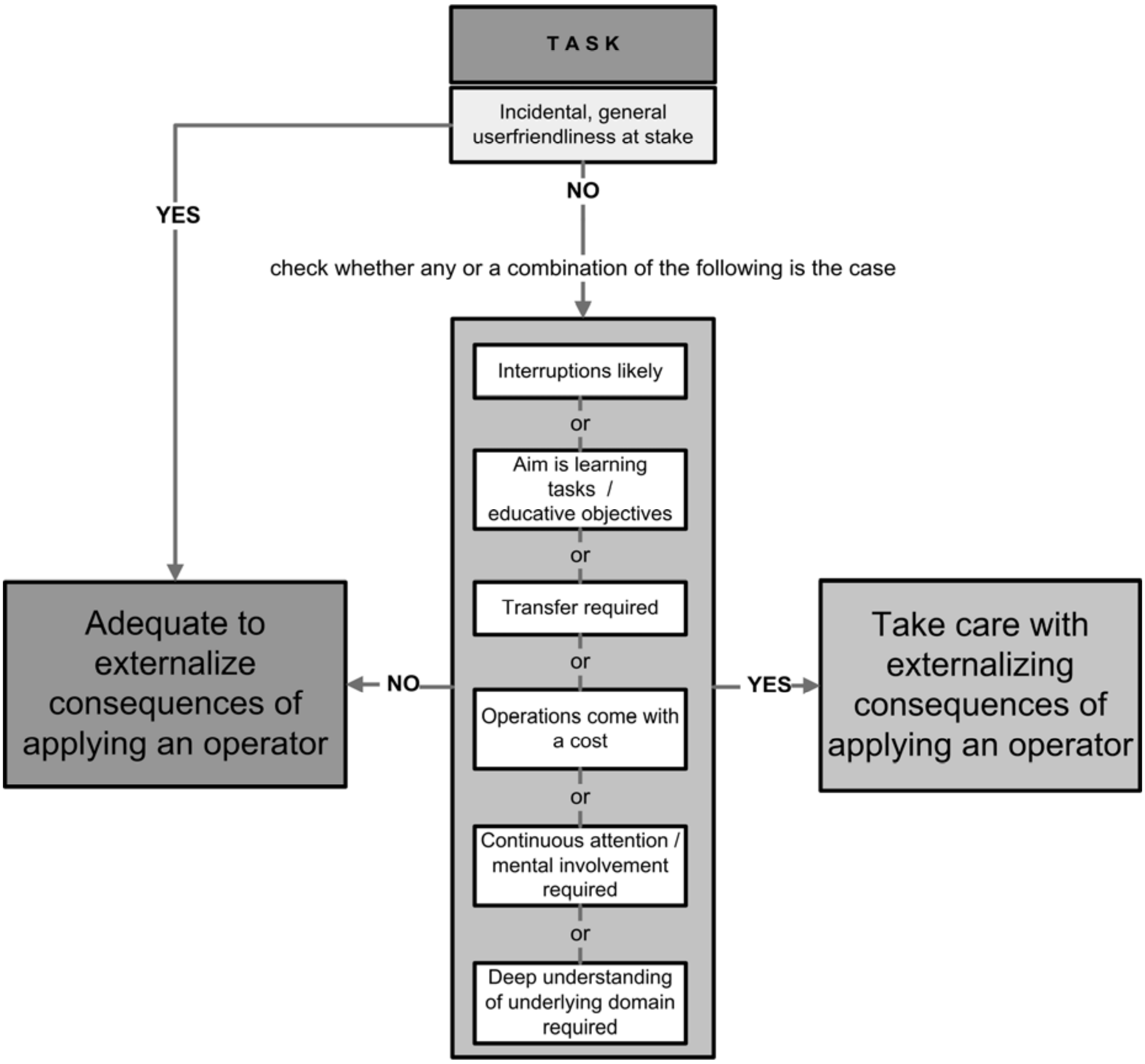



Figure 6.2 – Interaction framework

## 6.4 An excursion: Our theoretical notions and concepts applied to Computer Games

We end this chapter by devoting some special attention to the issue of computer games. Our experiments that took place in a HCI context showed that minimizing the information that interface operators convey could have a positive influence on the planfulness of behavior, focus, and attention, and on the levels of processing of the user, since the user is provoked to do more thinking himself instead of relying on the interface. There is game research by Kiili (2006), in which he emphasized the importance of distinguishing between activities related to solving the tasks, and the use of the controls (the interface, which he calls "artifact") of an educational game. He stated that "*all possible resources should be available for relevant processing (the main task) rather than for game control issues*". This is reminiscent of the primary and secondary tasks in the way we approached it and our statement that one has to be careful in letting the interface controls be part of the primary task. It is interesting to look at computer games in the context of our research, since they are extreme examples of software environments in which interactions, or reaching a goal is purposely made hard much of the time. Game designers deliver products that are able to cling users to their game consoles and quite seldom one sees such a determination and engagement as when watching gamers playing games, going through tremendous efforts to reach their goals. From computer games, we can take the step to the realm of educational games, serious games, training, and simulations. Regarding the latter, an experiment from Waldron, Patrick, Morgan and King (2007) showed that increased accessibility to externalized information had unwanted consequences. In a flight simulation, they found that memory for fused information on a display improved, when the availability the information within the interface was reduced. They concluded that one has to carefully watch the sensitive balance that exists between reducing cognitive workload (by making information more accessible) and ensuring that extraction of information is not made so effortless that information is not learned and is promptly forgotten. "Serious games" are software applications developed with game technology and game design principles for a primary purpose other than pure entertainment. Serious games are seen as similar to educational games, but are often intended for an audience outside of primary or secondary education. Games are famous for their motivational or "fun" aspects, but have the potential to facilitate cognitive processes such as making inferences and deeper thinking, which can be beneficial in an educational context.

The possible connection between games and our research is *bi-directional*. We could take advantage from what we can learn from the game industry concerning the way they keep players motivated and provoke continuous effort (Burgos, Van Nimwegen, Van Oostendorp & Koper, 2008). Likewise, there are features in games, which are reminiscent of what we have examined in our experiments. Our research issues could be valuable in decisions of game designers concerning interactivity and the challenge of keeping users motivated, proactive, and ready to make inferences, and not rely on the interface.

### 6.4.1 Information gaps

Motivation is defined as the arousal, direction, and persistence of behavior (Franken, 1994); an internal state or condition that activates behavior and gives it direction; a desire or want that energizes and directs goal-oriented behavior; it regulates the intensity of behavior. Central to many games in any case is that curiosity is being fostered and that there is something to discover, something has to be achieved, and not all is given away too easily. What was the defining quality of a game such as Myst (1993)? Central to its success was the fact that the player does not know anything, and has to solve a great number of riddles and questions on his own, the user is constantly triggered by *not knowing* things, and having to uncover them is a reason why the game is so intrinsically motivating. Regarding *not knowing things,* we will elaborate on the notion of *knowledge gaps*, or *information* gaps. Regarding curiosity and motivation, there is research in which cognitive and information-processing factors have been used to explain curiosity. Regarding the discrepancy between what one knows, and what there is to know (which information is available), an interesting perspective to look at motivation is the notion of knowledge gaps, also referred to as information gaps (Loewenstein, 1994). A knowledge gap refers to the difference between what a person knows and what is presented to a person. When one becomes aware of the knowledge gap, curiosity arises (Berlyne, 1954, 1960). Then, the awareness of the information gap produces an aversive feeling of deprivation or discomfort that can be alleviated only by obtaining the information needed to close the gap, which consequently produces an intense desire to modify the existing knowledge structure (Berlyne, 1960). This fits well with the general notion that learning is more effective when people experiment and discover for themselves. Exploratory learning as a whole has been a subject of research in many domains. Carroll (1990) for example, almost two decades ago already propagated minimalism in design and instructions. An example is to leave instructions intentionally incomplete to promote intrinsic interest through a problem challenge. In the context of “not knowing things” as in information gaps, and “wanting to know” there is an interesting experiment by Wishart (1990), who used a simulation to teach children what to do in case of a domestic fire. It showed that the amount of learning from an educational computer game increased by making it more challenging (among other mechanisms). Subjects learned more from a simulation that had combination of challenge and complexity. She concluded that educational software could be improved by providing a greater variety of challenge and complexity with less emphasis on reinforcement. Information gaps are used as a technique to “move students out of their seats”, allowing critical thinking and opinion formation. These processes are equivalent of the Germane Cognitive Load component of CLT in that it is induced by the learner's effort to process and comprehend the provided information. Realizing that information is missing makes the user contemplate and ask specifically about that information.

Regarding the issue of motivation and workload, within CLT, motivation for a task resides in the Germane Cognitive Load component, which when increased, is beneficial for learning if the mental effort does not exceed available processing capacity. To permit Germane Cognitive Load to be high, Extraneous Cognitive Load should be kept as low as possible. Proactive engagement, planfulness, and motivation, exactly what we have tried to provoke in our experiments, are key defining features of successful computer games. In the context of educational games and game based learning, several interactive learning techniques are used. Some examples are learning from mistakes, goal-oriented learning, role-playing, and constructivist learning. Engagement, motivation, and educational goals can mutually support each other in the same environment to achieve specific targets, i.e., learning content, improving personal and social skills, and working on strategies.

To describe the events in our experiments in terms of information gaps, it could very well be that users in the Externalization versions did not experience an information gap, they had the feeling that corresponds with the opposite, the notion that “knowledge is complete” because of the guidance they received. Perhaps the destination feedback gave them the feeling that solving the problem was nothing more than placing speakers following the green destination feedback (externalized information). In the Internalization version, however, subjects were not provided with the destination feedback, and more effort had to be invested in internal processes. They were aware that there was something to be done in order to solve the problem in a smart manner: to invest some effort, think a bit deeper, and develop an adequate strategy. We can think of two scenarios of what happens in the situations depicted in Figure 6.3.

**1. In the case of an Internalization interface (continuous arrow in Figure 6.3).**
The user is working with a computer application. In the case of the Internalization interface, the user is confronted with missing information as compared with the default situations in software that would be (at least at face value) beneficial to solve the problem. The user is lacking information, has to work in a plan-based fashion by coming up with a strategy. Inferences have to be made, which in this case means that the user has to start thinking for himself (the Internalization interface provokes this).

**2. In the case of an Externalization interface (dashed arrows in Figure 6.3).**
The user is working with a computer application. With the Externalization interface, the user is presented with information regarding the legality of moves; all the possible actions are highlighted. This information captures the user's attention, and will not so much seduce the users into mindful processes, because the user feels that the work is partly being done by the system. The user works in a display-based manner in that he decides actions based on interface cues to a high degree.

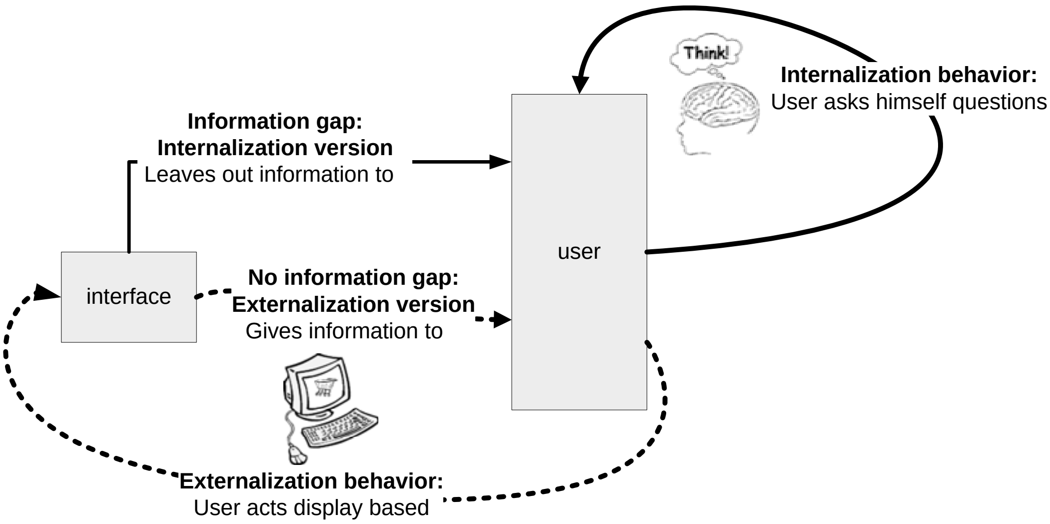



Figure 6.3 – Information gaps: Internalization and Externalization

We think the conception of information gaps and the mechanisms associated with them can be an interesting way to position our concepts of Internalization and Externalization in. Creating an information gap would correspond with what we called Internalization whereas *not* creating an information gap corresponds with what happens with Externalization.

### 6.4.2 Destination feedback in relation to informational gaps

Since what we have been varying in our experiments is how externalized information, in this case destination feedback (section 1.4) influences motivation and consequently plan-based behavior, it is worthwhile to look deeper into the issue of feedback in computer systems and its influence on behavior. As said, destination feedback is more or less default in a wide range of applications today, but in the preceding chapters, we saw that it does not necessarily improve performance better (among other findings). On the contrary, we saw that *not* doing this, thus creating an informational gap could trigger contemplation and effort from users.

This notion does not concur with some general notions stating that feedback in the traditional sense of the word is critical for learning since it provides support on the educational process and motivation. Also in the context of games, several authors stress the importance of feedback. They state that specific, contextual, and instant feedback based on goal commitment increases the motivation and effort, and consequently the performance of the learner. They also advocate the use of feedback to support game-based learning as a way to provide the learner with useful and immediate information about his performance. In the definition of Mason and Bruning (1999), feedback is defined as any message generated in response to a learner's action, usually *after* something is done. It implies that there is an interactive flow between the user and the system, i.e. that it involves processes in which information goes back and forth between a system and a user. Furthermore, this information flow is seen as a series of frequent inputs and not as a single one, because it is a part of the entire learning flow (Pivec, Dziabenko & Schinnerl, 2003). Through appropriate feedback, the learner is able to receive some information concerning the way he acts and learns. What follows can be seen as metacognitive activity, since the user is triggered to assess his own progress regarding his goals and actions, and he is able to make a consequent choice about the next action to take or even about the strategy to follow.

However, not all authors agree on the positive effects of feedback. After a literature review Mitchell and Savill-Smith (2004) state that the feedback mechanism could be perceived as a learning inhibitor, making the learner focuses on the outcome more than on analysis and strategy. They refer to a study by Halttunen and Sormunen (2000) that used an educational game to support learning of information retrieval. After an evaluation, the effectiveness feedback users received from the system was generally seen to promote learning significantly. Feedback concerning the performance of one's own query and the chance to reformulate the query and to further evaluate the effect of changes on performance was seen as a highly motivating and learning advancement. However, there were also students reporting that their attention was fixed on performance and they tried to improve their results mechanically, without analysis and reflection of their preceding queries and results, which points at little plan-based behavior. Here, the feedback tempted searchers to pay attention to the performance measures achieved, rather than on analysis of the search task situation and strategy. This is reminiscent of our assumptions. From HCI research, Gentner and Nielsen (1996) wrote an influential article called "The Anti-Mac Interface", not meant to discard the Macintosh interface guidelines, but to explore alternative approaches to computer interfaces. One of the issues here was to consider bringing some flexibility into the feedback and dialogue provided by the system. They pondered that rather than always providing the user with feedback on activities, the computer should be more flexible in the amount of feedback it provides. Initially, the computer could provide detailed feedback to familiarize the user with operations and instill confidence. Later, the feedback could be scaled back over time and restricted to unusual circumstances or times when the user requests more feedback.

The notion of informational gaps, for example by purposely *not* providing users with destination feedback (feedback that might provide users with the feeling of guidance) can be worthwhile with regard to game-based learning. Deliberately *not* providing guidance or assistance in this way and thus allowing Germane Cognitive Load so that learners can devote attention to development of proper strategies can be promising in also this context.

# 7 Conclusion

Nowadays we are used to friendly systems in which usability principles are implemented, and that are designed in such a way that a broad range of users can efficiently use them. Users are provided with visual feedback, guidance, and systems allow for errors (undoing actions is a common option). This causes users to rely on the interface and the guidance that can be derived from it, and interactions are characterized by *display-based* behavior. In the past, the knowledge to work with a computer came from the memory of a user, who was in constant command of a large body of knowledge and worked in a *plan-based* manner (O'Hara & Payne, 1999; Vera & Simon, 1993). Regarding guidance and feedback, we focus on the issue of the *Externalization* of information. By *Externalization* is meant that the necessity to remember certain characteristics of operators is eliminated or lessened, e.g. by limits on the applicability of operators. A way to implement this is to make parts of an interface context-sensitive, e.g. by hiding or disabling functions that are inapplicable at that moment. When no such features are provided, a user has to *internalize* the information himself and store it in his memory.

We looked on a deeper level what these presumed friendly systems actually do behavior, and questioned whether the highest achievable usability possible in fact is *always* desirable, since not all interactive systems have *only* the goal of "ease of use" and "maximum user satisfaction". There are situations where 100% command of functions and continuous proactive thought and planfulness are essential. Can it be fruitful to sometimes *not* make a system as assisting and guiding as one is used to? We wondered whether the Externalization of information by *destination feedback* might be a potential risk, because users could think they do not need to reason for themselves. It might seduce them to neglect to look for underlying rules and make inferences. And if this is the case, does this have a repercussion on performance, memory of the task and underlying rules, the solidness with which it is imprinted in memory, transfer of skill and task resumption after a severe interruption?

There is literature advocating Externalization that has shown that externalized information can be beneficial in serving the user. It can result in better performance (Zhang & Norman, 1994) because it made tasks easier. However, there are also findings stating that making an interface harder (on the operator level) results in better performance (O'Hara & Payne, 1998; 1999) because more plan-based behavior takes place. If one thinks in general terms about the opposition between "making it easier" (alter representation for better performance) or "making it harder" (alter operator/controls for better performance) it could seem that the findings of mentioned authors are contradicting. However, if one scrutinizes more, it becomes clear that the two currents of research (alter representation vs. alter operator/controls) are not contradicting each other on the same dimension. We connected these currents and positioned our angle on the matter as being *between* these two. We used *destination feedback* as a means of Externalization. It can guide decisions in showing which directions are possible. Externalization of information by destination feedback has been widely advocated by authorities in providing guidelines (Sun, 2001; Apple, 1992) and is default in modern computer interfaces. What is actually varied how operators (part of the secondary task, the interface) become part of the primary task by *dynamically* displaying information. As said, in our research, we did not alter the representation or the operators per se, but a combination of these two; we varied the external representation *via the operator*, *the* controls in the interface become the conveyors of externalized information. The Externalized information shows what objects can do *at this moment*; it externalizes the *consequences and enforcement* of rules. When a user drags an item from its place to a destination, the application provides feedback indicating whether it will accept that item. This can feel like guidance or assistance making the task easier, because less inferences and computation have to be performed by users.

Two series of experiments were conducted using two interface styles: one version in which certain information is externalized onto the interface (Externalization) and another version where this is not done (Internalization). The first series of experiments used a puzzle task, and the second series experiments used a more realistic task. Our hypotheses assumed that an Internalization interface, which may seem less friendly than could be, could be beneficial for task performance. We assumed that purposely *not* providing users with externalized information by providing destination feedback (perceived as assisting and guiding) has several beneficial effects. Externalization provides users with the feeling that the work is being partly done for them, and this could discourage active contemplation and investment of effort in the task. We will draw our conclusions analogue to the three levels of research questions we started with, namely a behavioral, a theoretical, and a practical level.

## 7.1 Conclusions on the behavioral level

We are now able to answer our specific research questions (section 1.5.1)

1. In repeated problem solving tasks, how does problem solving performance evolve over time?
   In the puzzle tasks, Internalization subjects tended to solve more trials, and reached fewer dead-end-states and superfluous moves after a small delay. Externalization subjects did not perform the tasks faster, and had worse declarative knowledge of the problem and its solutions afterwards. Also in the more realistic task there were no *overall* time differences, but Internalization subjects took more time to think before they started, and between the individual moves. Internalization subjects performed less superfluous moves, applied smarter strategies, and again had better knowledge afterwards. In the beginning of the tasks, Externalization subjects spent significantly more time looking at the destination feedback, which took place in a part of the screen where the most relevant information could *not* be found, indicating display-based behavior.
2. In the case of transfer of skill to another task, does learning and doing a task in one interface style have an influence on performance on a similar (but different) task?
   In the realistic task it showed that performance on a similar task was better if the initial task had been done using the Internalization interface. The Internalization interface causes better imprinting of knowledge and that solutions are more ready available and applicable in a transfer situation.
3. With transfer of skill to another interface style, does learning and doing tasks in one interface style have an influence performing that same task afterwards with the other interface style?

In the realistic task, Internalization subjects had better transfer of skill to another interface. Switching *interface* in the sequence INT→EXT resulted in better performance than EXT→INT.

4. What is the influence of instruction to the user before the tasks (instruction to plan carefully, vs. telling users that mistakes are not a problem) on performance with the two interface versions?
   When an instruction to plan carefully was presented, only the Internalization subjects were influenced by it, they invested more effort in finding a solution and performed better.
5. Do individual differences in cognitive style play a role in this?
   Preexisting cognitive style of users was measured by the Need for Cognition scale, reflecting the tendency to engage in, and enjoy effortful cognitive tasks. It had no influence on behavior and performance; it was our manipulation only that caused differences in performance.
6. Delay: do the two interface styles have different effects on performance after a large delay?
   With the puzzle task, after a large delay, Internalization subjects had better memory for the same task, and performed better (also on a transfer task).
7. Do interruptions of the problem solving process have different effects in different interface styles?
   In the realistic task, Internalization subjects were better at task resumption after a severe interruption. Internalization subjects continued as if nothing happened, whereas the performance of the Externalization subjects deteriorated badly.

Resuming, it showed that immediate and delayed performance was worse for users using the Externalization interface. In addition, transfer of skill was worse for these users, both to another task, and to another interface. Internalization subjects imprinted relevant task and rule knowledge better and were not affected by a severe interruption in the workflow, whereas Externalization subjects were.

## 7.2 Conclusions on the theoretical level

By answering the above questions, we obtained more insight in how plan-based vs. display-based problem behavior is provoked by the Internalization respectively Externalization interface style (see also chapter 6). We connected this to two concepts from Cognitive Load Theory. These concepts are Extraneous Cognitive Load, which results from the design of the material and refers to *unnecessary* load caused by inefficient design, and Germane Cognitive Load, which refers to the degree of effort involved in processing (effective for learning and is associated with motivation and interest). Based on our findings with the type of tasks and interfaces we used, we conclude the following:

**The Internalization interface**

- This interface is characterized by minimal intertwining of the primary task (solve the problem) and the secondary task (manage controls, the interface). No mental steps are taken over by the interface; the tool (interface) stays a tool, and does not provide the

user with the feeling of assistance. As a result, strategy construction and inferences come from the users mind.
- The Internalization interface causes low Extraneous Cognitive Load. No resources are (unnecessarily) taken away from the task and users are not distracted by interface events, no Working Memory resources have to be devoted to processing them.
- Also a cost-benefit consideration plays a part; Internalization subjects realize quickly that aimlessly clicking around will not help. Plan-based behavior is provoked, triggering users to invest more focused effort in the task. This corresponds with high Germane Cognitive Load (directed at beneficial mental processes and associated with motivation and interest). Internalization subjects study the situation better, are less distracted, and follow a more solid plan.
- The result of the fact that Internalization subjects make more inferences and applied smarter strategies, is less trial and error problem solving, resulting in more straightforward solutions (efficiency).
- Doing the mental processing more by oneself causes better imprinted knowledge and insight immediately, and also after a large delay.
- Because users have internalized information and constructed strategies oneself, knowledge is also better applicable to similar problems in transfer situations.

**The Externalization Interface**
- Might allow cognitive offloading, but this in itself does not cause users to devote more effort to beneficial mental processes.
- The destination feedback provided users with the feeling that they were guided, and that the task is partially carried our for them. This discourages active contemplation, it makes users less willing to invest effort and make inferences. They behave lazier and shallower, Germane Cognitive Load is low, reflection is not encouraged and more display-based behavior occurs. This also overrides personal cognitive styles of users.
- Besides this phenomenon, the destination feedback still has to be processed by the user, and can be regarded as Extraneous Cognitive Load. It brings nothing positive; it acts as a distracter, and inhibits the focus on the problem itself. This was emphasized by Externalization subjects not paying any attention to planning instruction beforehand.
- When users depend to a high degree on the externalized feedback, behavior is largely display-based, accompanied by unnecessary actions that steer away from a direct solution.

Resuming, users that internalize information themselves behave more plan-based, are more proactive and ready to make inferences. This in turn, results in more focus, more direct and economical solutions, better strategies, and better imprinting of knowledge. This knowledge is easier to recall at a future point in time, and is better transferable to transfer situations where the interface, the task, or both were different, less vulnerable to a severe interruption, and better applicable to transfer situations.

We conclude with some general thoughts on our theoretical framework. We see Information Processing Psychology, which describes cognitive processes and claims that humans operate as an information processing system as too rational, assuming obedient problems solvers in isolated situations, and as such too limited for our HCI context. Task-execution in tasks as the ones we study is influenced by environmental factors (Suchman, 1987), the tool to do the task (the interface and its controls) and habits and expectations computer users have developed. All this has consequences for motivational and metacognitive factors that play a role during interactions, and this deserves attention.

We have – and we are not alone in this - applied Cognitive Load Theory (CLT), which provides guidelines to assist in the presentation of information in a manner that encourages learner activities that optimize performance and is concerned with limitations of Working Memory. CLT has the basic premise that learning is hindered if the instructional material overwhelms learners' cognitive resources, which is the case in the Externalization interface. Regarding the same CLT, we criticized one of the main assumptions stating that cognitive load in general should be low. It depends on the kind of cognitive load. Along the way and after accumulating thoughts and insights, we conclude that CLT is valuable, but on its own cannot account for the complete range of processes that takes place, just as IPP cannot. Concepts such as Extraneous and Germane Cognitive Load are valuable in describing events that happen in the periphery of problem solving in computer-based tasks, such as how users behave in certain circumstances, and how interface-related events can change our behavior in approaching a task, especially concerning the motivation and effort involved. However, the actual informational processes that problem solvers engage in would be incomplete without notions from IPP. Alongside the notion of cost-benefit, we see Cognitive Load as influencing Information Processing processes.

## 7.3 Conclusions on the practical level

Regarding the empirical findings and our conclusions, we state that there is more to HCI guidelines than how they are now. Our findings show that computer-mediated tasks can take advantage of interfaces that are designed from considerations that run deeper than plain usability, even when they go against common sense. We do not credit ourselves with the thought itself, since we were inspired by preceding research that paved the way. However, our experimental findings revealed interesting results regarding a specific kind of Externalization, which is very common: destination feedback. We concluded that relieving a user's memory and making interactions assisted by externalizing information does not have beneficial effects. It makes users count on the interface and gives them (unrightfully so) the feeling that the task and thinking-work is partly done for them, which seduces users in more shallow cognitive behavior.

There are of course situations in which this is not crucial at all. However, depending on the specific goal and situation, it can be valuable to design software in such a way that making inferences is provoked, and active learning takes place and users have better command of what it is they are doing. It can be important when dependence on a particular interface is specifically undesired, such as situations where risky and complex tasks are performed, and where a user suddenly is confronted with a new situation. Now, his/her insights and knowledge has to be transferred to a different situation, and this requires stored plans. The common guideline to "not give users the chance to make mistakes" should of course not be neglected, but at the same time, interaction should facilitate or even persuade users to learn what underlies the task they are doing. The same is true in situations where interruptions are commonplace and where in the meanwhile mastery of what is underlying a task or domain is desired, or when operations come with a cost and direct solutions without deviations are the aim.

In designing our interfaces we have to be careful with providing interface cues that give away too much, and must design in such a manner that the way users (should) think is optimally supported, which in turn could help the software to achieve its specific goal. We therefore presented an interaction framework, which can guide decisions in specific situations that have characteristics that go beyond plain usability, without presuming our framework is complete regarding all other tasks, software, or environments. It can prove value and guide decisions regarding interface design and/or interactive properties of the interface in more serious contexts, for example educational contexts where learning itself is the aim or safety critical task situations.

## 7.4 Future directions

There are various directions we could think of to extend our research. There are more ways to externalize information, and we are aware that Externalization and Internalization by no means constitute a dichotomous variable, but should be positioned on a continuum. In our experiments, we never had a situation in which the Externalization was zero; we made certain choices. Neither did we have situations in which *even more* was externalized, as in Kotovsky et al.'s *all*-info version, which is interesting to look into, also in more realistic software applications.

Another thought is that we could vary the amount of Externalization for example with some adaptivity. It would be possible to adaptively reduce or increase Externalization during the process, or to do this depending on the displayed behavior of subjects.

An issue we have to mention is user expertise. We knew that none of the subjects knew Ball & Boxes or Conference Planner. In the tasks where transfer was studied, or in the session where performance after a large delay was studied, we could say our users all had some experience. However, on the more general level of computer experience, puzzle experience or experience with office-like administration task we did not focus. Since we mention users' habits and expectations as playing a role in behavior, it might be worthwhile to incorporate this in future research, since it might very well be that a "blank" person with almost no computer experience would display different behavior. Expertise in the sense of skill or knowledge in a particular domain also can be worthwhile to incorporate, especially in educational contexts. We only included the Need for Cognition scale, which refers to attitude, rather than expertise, and did not seem to influence performance in combination with our interface styles.

Regarding serious gaming, game based learning and training situations; we hope to partially have paved the way for more research in that direction coming from our background. Regarding Berlyne's (1960) curiosity principle, it might too pretentious to claim that leaving out information as we did would arouse *an epistemic feeling of discontent not to know things*. Nevertheless, an information gap can play a part, albeit less drastically, and can be fruitful for these research areas. The more gaming moves towards educational realms, training and the like, the more interesting it can be to see how attention, mental effort and proactive behavior can be safeguarded. Likewise, the areas of educational systems and training can learn from principles that make games so successful.

A last word should be devoted to the tasks we used. We deliberately started with a puzzle-like, highly controllable task, and applied and implemented Externalization features to it. Next, we used more realistic software that had the physical properties and behaviors users are used to nowadays. Correctness in terms of whether the tasks were successfully performed or not, was not so much the focus, since we wanted to analyze solution paths and strategy when the tasks were solved. Although a justified comment might be that our findings were obtained in a certain category of tasks, that various issues have not been taken in account for, and that there is a lot to explore still (which we acknowledge), we still claim that our findings add to the body of theory regarding problem solving and interaction in computer-based tasks. It was our choice to use tasks that were of medium difficulty. Next steps in future research could incorporate more complex tasks in which the intrinsic load of the tasks itself is very high, and also include multi-tasking situations, or even groupware situations.

## Bibliography

Anderson, J. R. (1990). *The adaptive character of thought*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Apple Computer, Inc. (1987). *Human Interface Guidelines: The Apple Desktop Interface*. Reading, MA: Addison-Wesley Publishing Co.

Apple Computer, Inc. (1992). *Macintosh Human Interface Guidelines*. Boston, MA: Addison-Wesley Publishing Co.

Arroyo, E., Selker, T., & Wei, W. (2006). Usability Tool for Analysis of Web Designs Using Mouse Tracks. In Proceedings of CHI'06: *The SIGCHI Conference on Human Factors in Computing Systems*. New York, NY: ACM.

Back, J., & Oppenheim, C. (2001). A model of cognitive load for IR: implications for user relevance feedback interaction. *Information Research,6*(2). Retrieved December 11, 2007 from http://informationr.net/ir/6-2/ws2.html.

Bannon, L. J. (1991). The Role of Psychology and Human-Computer Interaction Studies in Systems Design. In J. Greenbaum & M. Kyng (Eds.), *Design at work: Cooperative Design of Computer Systems* (pp. 25-44). Mahwah, NJ: Lawrence Erlbaum Associates.

Berlyne, D. E. (1954). A theory of human curiosity. *British Journal of Psychology, 45*(3), 180-191.

Berlyne, D. E. (1960). *Conflict, Arousal, and Curiosity*. New York, NY: McGraw Hill.

Boucheix, J. M. (2004). Multimedia Simulation and Comprehension Aid for Complex Technical Documents: A Course for Crane Drivers. Proceedings of ICALT '04: *The IEEE International Conference on Advanced Learning Technologies.* Washington, DC: IEEE Computer Society.

Brusilovsky, P. (2001). Adaptive hypermedia. *User Modeling and User-Adapted Interaction, 11*(1-2), 87-110.

Burgos, D., Van Nimwegen, C., Van Oostendorp, H., & Koper, R. (2008). Game-based learning and the role of feedback. A case study of The Planning Educational Task. *International Journal of Advanced Technology for Learning (in press).*

Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology, 42*(1), 116-131.

Cacioppo, J. T., Petty, R. E., & Kao, C. E. (1984). The efficient assessment of need for cognition. *Journal of Personality Assessment, 48*(3), 306-307.

Cacioppo, J. T., Petty, R. E., Feinstein, J., & Jarvis, W. B. G. (1996). Dispositional differences in cognitive motivation: The life and times of individuals varying in need for cognition. *Psychological Bulletin, 119*(2), 197-253.

Card, S. K., Moran, T. P., & Newell, A. (1983). *The Psychology of Human-Computer Interaction.* Mahwah, NJ: Lawrence Erlbaum Associates.

Carroll, J. M. (1990). The Nurnberg Funnel: Designing Minimalist Instruction for Practical Computer Skill. Cambridge, MA: MIT Press.

Carroll, J. M., & Rosson, M. B. (1987). The paradox of the active user. In J. M. Carroll (Ed.), *Interfacing Thought: Cognitive Aspects of Human-Computer Interaction* (pp. 80-111). Cambridge, MA: MIT Press.

Chen, M., Anderson, J. R., & Sohn, M. (2003). Eye-Hand Coordination During Web Browsing. In J. Ratner, J. (Ed.), *Human Factors and Web Development* (pp. 207-224). Mahwah, NJ: Lawrence Erlbaum Associates.

Chen, C., & Rada, R. (1996). Interacting with hypertext: a meta-analysis of experimental studies. *Human-Computer Interaction*, *11*(2), 125-156.

Crystal, A., & Kalyanaraman, S. (2005). Personality variables and information processing on the Web: methodological issues and empirical evidence. Paper presented at: *The International Communication Association Annual Meeting*, New York, NY, May 26 - 30.

Ericsson, K. A., & Kintsch, W. (1995). Long-Term Working Memory. *Psychological Review, 102,* 211-245.

Ernst, G. W., & Newell, A. (1969). *GPS: A Case Study in Generality and Problem Solving*. New York, NY: Academic Press.

Feinberg, S., & Murphy, M. (2000). Applying cognitive load theory to the design of web-based instruction. Proceedings of IPCC/SIGDOC '00: *IEEE professional communication society international professional communication conference.* Piscataway, NJ: IEEE Educational Activities Dept.

Franken, R. E. (1994). *Human Motivation (3rd ed.).* Belmont, CA: Brooks/Cole Publishing.

Fritz 8 (2003). Morsch, F., & Feist, M., ChessBase GmbH.

Fu, W., & Gray, W. D. (2004). Resolving the paradox of the active user: stable suboptimal performance in interactive tasks. *Cognitive Science, 28*(6), 901-935.

Gentner, D., & Nielsen, J. (1996). The Anti-Mac Interface. *Communications of the ACM*, *39*(8), 70-82.

Gerjets P., & Hesse F. (2004). When are powerful learning environments effective? The role of learner activities and of students' conceptions of educational technology. *International Journal of Educational Research, 41*(6), 445-465.

Gerjets, P., & Scheiter, K. (2003). Goal configurations and processing strategies as mediators between instructional design and cognitive load: Evidence from hypertext-based instruction. *Educational Psychologist, 38*(1), 33-41.

Gerjets, P., Scheiter, K., & Catrambone, R. (2004). Designing instructional examples to reduce intrinsic cognitive load: Molar versus modular presentation of solution procedures. *Instructional Science, 32(1-2),* 33-58.

Gick, S., & Holyoak G. (1983). Schema Induction and Analogical Transfer. *Cognitive Psychology, 15*(1), 1-38, 1983.

Gould, J. D., & Lewis, C. (1983). Designing for usability: Key principles and what designers think. *Communications of the ACM*, *28*(3), 300-311.

Gray, W. D., & Fu, W. T. (2001). Ignoring perfect knowledge in-the-world for imperfect knowledge in-the-head. In Proceedings of CHI '01: *SIGCHI Conference on Human Factors in Computing Systems.* New York, NY: ACM.

Guttormsen, S., Schär, S., Schluep, S., Schierz, C., & Krueger, H. (2000). Interaction for computer-aided learning. *Interactive Multimedia Electronic Journal of Computer-Enhanced Learning*, *2*(1). Retrieved December 04, 2007, from http://imej.wfu.edu/articles/2000/1/03/index.asp.

Halttunen, K., & Sormunen, E. (2000). Learning information retrieval through an educational game. Is gaming sufficient for learning? *Education for Information*, *18*(4), 289–311.

Hayes, N. A., & Broadbent, D. E. (1988). Two modes of learning for interactive tasks. *Cognition, 28*(3), 249-276.

Hollan, J., Hutchins, E., & Kirsh, D. (2000). Distributed cognition: toward a new foundation for human-computer interaction research. *ACM Transactions on Computer-Human Interaction, 7*(2), 174-196.

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

Inhoff, A. W., & Radach, R. (1998). Definition and computation of oculomotor measures in the study of cognitive processes. In G. Underwood (Ed.), *Eye guidance in reading, driving and scene perception* (pp. 29-53). New York, NY: Elsevier.

Jacob, R. J. K., & Karn, K. S. (2003). Eye tracking in human-computer interaction and usability research: Ready to deliver the promises. In J. Hyönä, R. Radach & H. Deubel (Eds.), *The Mind's Eye: Cognitive and Applied Aspects of Eye Movement Research* (pp. 573-605). Amsterdam, The Netherlands: Elsevier Science.

Jeffries, R., Polson, P., Razran, L., & Atwood, M. (1977). A Process Model for Missionaries-Cannibals and Other River-Crossing Problems. *Cognitive Psychology, 9*(4), 412-440.

Kiili, K. (2006). Evaluations of an experiential gaming model. *Human Technology, 2*(2), 187–201.

Kirschner, P. A. (2002). Cognitive load theory: implications of cognitive load theory on the design of learning. *Learning and Instruction, 12*(1), 1-10.

Koppel, R., Metlay. J. P., Cohen, A., Abaluck, B., Localio, A. R., Kimmel, S. E., & Strom B. L. (2005). The Role of Computerized Physician Order Entry Systems in Facilitating Medication Errors. *Journal of the American Medical Association, 293*(10), 1197-1203.

Kort, B., Reilly, R., & Picard, R. (2001). External Representation of Learning Process and Domain Knowledge: Affective State as a Determinate of its Structure and Function. Proceedings of *The IEEE Artificial Intelligence in Education Conference*. Amsterdam, The Netherlands: IOS Press.

Kotovsky, K., Hayes, J. R., & Simon, H. A. (1985). Why are some problems hard? Evidence from the Tower of Hanoi. *Cognitive Psychology, 17*(2), 248-294.

Kotovsky, K., & Simon, H. A. (1990). What makes some problems really hard: Explorations in the problem space of difficulty. *Cognitive Psychology, 22*(2), 143-183.

Landauer, T. K. (1996). *The Trouble with Computers: Usefulness, Usability, and Productivity*. Cambridge, MA: MIT Press.

Larkin, J. H. (1989). Display-based problem solving. Complex information processing. In D. K. S. Kotovsky (Ed.), *The Impact of Herb Simon* (pp. 319-332). Hillsdale, NJ: Lawrence Erlbaum Associates.

Larkin, J. H., & Simon, H. A. (1987). Why a Diagram is (Sometimes) Worth Ten Thousand Words. *Cognitive Science, 11*(1), 65-100.

Loewenstein, G. (1994). The psychology of curiosity: a review and reinterpretation. *Psychological Bulletin,* 116(1), 75-98.

Mason, B. J., & Bruning, R. (1999). *Providing Feedback in Computer-based Instruction: What the Research Tells Us.* Retrieved December 04, 2007 from http://dwb.unl.edu/Edit/MB/MasonBruning.html.

Mayer, R. E. (1983). *Thinking, problem solving, cognition*. New York, NY: Freeman.

Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist, 38*(1), 43-52.

Mayes, J. T., Draper, S.W., McGregor, A. M., & Oatley, K. (1988). Information flow in a user interface: the effect of experience and context on the recall of MacWrite screens. In D. M. J. R. Winder (Ed.), *People and Computers* (pp. 275 - 289). Cambridge, MA: Cambridge University Press.

Miller, G. A., Galanter, E., & Pribram. K. H. (1960). *Plans and the structure of behavior*. Holt, NY: Rinehart and Winston.

Mitchell, A., & Savill-Smith, C. (2004). *The use of computer and video games for learning. A review of the literature.* Retrieved December 13, 2007 from http://www.lsda.org.uk/files/PDF/1529.pdf.

Moreno, R., & Mayer, R. E. (2000). Meaningful Design for Meaningful Learning: Applying Cognitive Theory to Multimedia Explanations. In Proceedings from EDMEDIA '00 : *World Conference on Educational Multimedia, Hypermedia and Telecommunications 2000* (pp. 782-787). Chesapeake, VA: AACE.

Myst (1993). Robyn Miller & Rand Miller, Brøderbund Software.

Nagata, S. (2006). *User Assistance with Interruptions on a Mobile Device*. (Doctoral Dissertation, Utrecht University, The Netherlands).

Neerincx, M. A., & Griffioen, E. (1996). Cognitive task analysis: harmonizing tasks to human capacities. *Ergonomics*, *39* (4), 543-561.

Neerincx, M. A., Pemberton, S., & Lindenberg, J. (1999). U-WISH Web usability: methods, guidelines and support interfaces. Soesterberg, The Netherlands: TNO Human Factors Research Institute.

Newell, A., & Simon, H. (1972). *Human problem solving.* Englewood Cliffs, NJ: Prentice-Hall.

Nielsen, J. (1993). *Usability Engineering*. San Francisco, CA: Morgan Kaufmann Publishers Inc.

Norman, D. A. (1988). *The Psychology of Everyday Things*. New York, NY: Basic Books.

O'Hara, K. P. (1994). Cost of Operations Affects Planfulness of Problem solving Behavior. In Proceedings of CHI '94: *Conference on Human Factors in Computing Systems.* New York, NY: ACM.

O'Hara, K. P., & Payne, S. J. (1998). The Effects of Operator Cost on Planfulness of Problem Solving and Learning. *Cognitive Psychology, 35*(1), 34-70.

O'Hara, K. P., & Payne, S. J. (1999). Planning the user Interface: the effects of lockout time and error recovery cost. *International Journal of Human-Computer Studies, 50*(1), 41-59.

Oulasvirta, A., & Saariluoma, P. (2006). Surviving task interruptions: Investigating the implications of long term working memory theory. *International Journal of Human Computer Interaction, 64*(10), 941-961.

Oviatt, S., Arthur, A., & Cohen, J. (2006). Quiet interfaces that help students think. In Proceedings of UIST'06: *The 19th Annual ACM Symposium on User Interface Software and Technology.* New York, NY: ACM.

Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive Load Theory and Instructional Design: Recent developments. *Educational Psychologist, 38*(1), 1-4.

Paas, F., Tuovinen, J.E., Van Merriënboer, J. J. G., & Darabi, A. A. (2005). A Motivational Perspective on the Relation Between Mental Effort and Performance: Optimizing

Learner Involvement in Instruction. *Educational Technology Research and Development, 53*(1), 24-34

Payne, S. J. (1991). Display-based action at the user interface. *International Journal Man-Machine Studies*, *35*(3), 275-289.

Perkins, D. N., & Salomon, G. (1992). Transfer of learning. *International Encyclopedia of Education* (2nd ed.). Oxford, UK: Pergamon Press.

Pieters, R. G. M., Verplanken, B., & Modde, J. M. (1987). Neiging tot nadenken: Samenhang met beredeneerd gedrag. *Nederlands Tijdschrift voor de Psychologie, 42*(1), 62-70.

Pivec, M., Dziabenko, O., & Schinnerl, I. (2003). Aspects of game-based learning. In Proceedings of I-KNOW '03: *The Third International Conference on Knowledge Management.* Berlin, Germany: Springer / KNOW-Center.

Plastelina Interactive LTD (2004). http://www.plastelina.net/.

Rasmussen, J. (1983). Skills, rules, and knowledge; Signals, signs, and symbols, and other distinctions in human performance models. *IEEE transactions on systems, man, and cybernetics, 13*(3)*,* 257-266.

Ridley, D. S., Schutz, P. A., Glanz, R. S., & Weinstein, C. E. (1992). Self-regulated learning: the interactive influence of metacognitive awareness and goal-setting. *Journal of Experimental Education 60*(4), 293-306.

Rikers, R. M. J. P., Van Gerven, P. W. M., & Schmidt, H. G. (2004). Cognitive Load Theory as a Tool for Expertise Development. *Instructional Science, 32* (1-2), 173-182.

Rubin, J. (1994). *Handbook of Usability Testing: How to Plan, Design, and Conduct Effective Tests*. New York, NY: Wiley.

Scaife, M., & Rogers, Y. (1996). External Cognition: how do graphical representations work? *International Journal of Human-Computer Studies*, *45*(2), 185-213.

Schnotz, W., & Kurschner, C. (2007). A Reconsideration of Cognitive Load Theory. Educational Psychology Review, *19*(4), 469-508.

Schnotz, W., & Rasch, T. (2005). Enabling, facilitating, and inhibiting effects of animations in multimedia learning: Why reduction of cognitive load can have negative results on learning. *Educational Technology, Research and Development*, *53*(3), 47-58.

Scott, B. M., & Schwartz, N. H. (2007). Navigational spatial displays: The role of metacognition as cognitive load. *Learning and Instruction, 17*(1)*,* 89–105.

Shackel, B. (1991). Usability - Context, Framework, Definition, Design and Evaluation. In B. Shackel & S. Richardson (Eds.), *Human Factors for Informatics Usability* (pp. 21-38). Cambridge, MA: Cambridge University Press.

Shneiderman, B. (1983). Direct manipulation: a step beyond programming languages. *IEEE Computer, 16*(8), 57-69.

Simpson, A., & McKnight, C. (1990). Navigation in hypertext: structural cues and mental maps. In R. McAleese & C. Green (Eds.), *Hypertext: State of the Art* (pp. 73-83). Oxford, England: Intellect.

Stanton, N. A., Taylor, R. G., & Tweedie, L. A. (1992). Maps as navigational aids in hypertext environments: An empirical evaluation. *Journal of Educational Multimedia and Hypermedia, 1*(4), 431-444.

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

Sun Microsystems, Inc. (2001). *Java Look and Feel Design Guidelines, second edition*. Cambridge, MA: Addison-Wesley Publishing Co.

Svendsen, G. (1991). The influence of interface style on problem solving. *International Journal Man-machine studies, 35*(3)*,* 379-397.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12*(1), 257-285.

Sweller, J. (1999). *Instructional design in technical areas*. Camberwell, Australia: ACER Press.

Sweller, J., Van Merriënboer, J. J. G., & Paas, F. (1998). Cognitive Architecture and Instructional Design. *Educational Psychology Review, 10*(3), 251-295.

Tabachneck-Schijf, H. J. M. (1992). Computational differences in mental representations: effects of mode of data presentation on reasoning and understanding. (Doctoral Dissertation, Carnegie-Mellon University, Pittsburgh, USA).

Todd, P. M., & Gigerenzer, G. (2000). Précis of Simple heuristics that make us smart. *Behavioral and Brain Sciences, 23*(5), 727-741.

Trudel, C. I., & Payne, S. J. (1995). Reflection and goal management in exploratory learning. *International Journal of Human-Computer Studies, 42*(3), 307-339.

Trudel, C. I., & Payne, S. J. (1996). Self-monitoring during exploration of an interactive device. *International Journal of Human-Computer Studies, 45*(6)*,* 723-747.

Unfold (2004). *The UNFOLD Project*, www.unfold-project.net.

Utting, K., & Yankelovich, N. (1998). Context and orientation in hypermedia networks. *ACM Transactions on Information Systems, 7*(1), 58-84.

Valcke, M. (2002). Cognitive load: updating the theory? *Learning and Instruction*, *11*(4), 147-154.

Van der Veer, G. C. (1990). Human-computer interaction: Learning, individual differences, and design recommendations (Doctoral dissertation, Vrije Universiteit Amsterdam, The Netherlands).

Van Nimwegen, C., & Burgos, D. (2005). Performance in a planning task: The (ir)relevance of interface style and user's cognitive style. Paper presented at SIGCHI.NL ‘05: *HCI Close 2U*. The Hague, The Netherlands, October 13.

Van Nimwegen, C., Burgos, D., Van Oostendorp, H., & Tabachneck-Schijf, H. J. M. (2006). The Paradox of the Assisted User: Guidance can be Counterproductive. In Proceedings of CHI’06: *The SIGCHI Conference on Human Factors in Computing Systems*. New York, NY: ACM.

Van Nimwegen, C., & Van Oostendorp, H. (2006). Visual guidance in the interface and mouse movement behavior. Do we look where the relevant information is? In Proceedings of InScit ’07: *Current Research in Information Sciences and Technologies*, (pp. 172-176). Merida, Spain: Open Institute of Knowledge.

Van Nimwegen, C., & Van Oostendorp, H. (2007a). Guidance in the interface and transfer of task performance. In proceedings of ECCE ’07: *The European Conference on Cognitive Ergonomics.* France: European Association of Cognitive Ergonomics.

Van Nimwegen, C., & Van Oostendorp, H. (2007b). Visual guidance and the impact of interruption during problem solving: Interface issues embedded in Cognitive Load Theory. In Proceedings of EuroCogsci ’07: *The Second European Cognitive Science Conference*. Mahwah, NJ: Lawrence Erlbaum Associates.

Van Nimwegen, C., & Van Oostendorp, H. (2008a). The questionable impact of an assisting interface on performance in transfer situations. *International Journal of Industrial Ergonomics* (under review).

Van Nimwegen, C., & Van Oostendorp, H. (2008b). Influencing Proactive User Behavior: Visual Assistance can be Counter-effective. *Interacting with Computers* (submitted).

Van Nimwegen, C., Van Oostendorp, H., Burgos, D., & Koper, R. (2006). Does an interface with less assistance provoke more thoughtful behavior? In Proceedings of ICLS '06: *The 7th international Conference on Learning Sciences*. Bloomington, IN: International Society of the Learning Sciences.

Van Nimwegen, C., Van Oostendorp, H., & Tabachneck-Schijf, H. J. M. (2004a). Can more help be worse? The over-assisting interface. In Proceedings of SIGCHI.NL '04: *Conference on Dutch directions in HCI*. New York, NY: ACM Press.

Van Nimwegen, C., Van Oostendorp, H., & Tabachneck-Schijf, H. J. M. (2004b). Does context sensitivity in the interface help? In proceedings of ECCE '04: *The Twelfth European Conference on Cognitive Ergonomics*. France: European Association of Cognitive Ergonomics.

Van Nimwegen, C., Van Oostendorp, H., & Tabachneck-Schijf, H. J. M. (2004c). Externalization vs. Internalization: The Influence on Problem Solving Performance. In Proceedings of ICALT '04: *The 4th IEEE Int. Conference on Advanced Learning Technologies*. Los Alamitos, CA: IEEE.

Van Nimwegen, C., Van Oostendorp, H., & Tabachneck-Schijf, H. J. M. (2004d). The over-assisting interface: The role of support in knowledge acquisition. In Proceedings of HCI2004: *Design for Life*. Bristol, UK: Research Press International (on behalf of the British HCI).

Van Nimwegen, C., Van Oostendorp, H., & Tabachneck-Schijf, H. J. M. (2005). The role of interface style in planning during problem solving. In Proceedings of CogSci '05*: The 27th Annual Cognitive Science Conference*. Mahwah, NJ: Lawrence Erlbaum.

Van Nimwegen, C., Van Oostendorp, H., & Tabachneck-Schijf, H. J. M. (2006). Guidance in the interface: Effects of externalizing information during problem solving. In E. M. Alkhalifa (Ed.), *Cognitively Informed Systems: Utilizing Practical Approaches to Enrich Information Presentation and Transfer* (pp. 74-101). Hershey, PA: Idea Group Inc.

Van Oostendorp, H., & De Mul, S. (1999). Learning by exploration: Thinking aloud while exploring an information system. *Instructional Science, 27*(3-4)*,* 269-284.

Van Oostendorp, H., & Beijersbergen, M. (2007). Animated diagrams: their effect on understanding, confidence and mental effort. Paper presented at EARLI: *The 12th Conference for Research on Learning and Instruction*. Budapest, Hungary, August 29- September 1.

Vera, A. H., & H. A. Simon (1993). Situated action: a symbolic interpretation. *Cognitive Science, 17*(1), 7-48.

Waldron, S. M., Patrick, J., Morgan, P. L., & King, S. (2007). Influencing Cognitive Strategy by Manipulating Information Access. *The Computer Journal, 50*(6), 694-702.

Williams, D. J., & Noyes, J. (2007). Effect of experience and mode of presentation on problem solving. *Computers in Human Behavior*, *23*(1), 258-274.

Wishart, J. (1990). Cognitive factors related to user involvement with computers and their effects upon learning from an educational computer game. *Computers Education, 15*(1-3), 145-150.

Zhang, J. (1993). External Representation: An Issue for Cognition. *Behavioral & Brain Sciences, 16*(4), 774-775.

Zhang, J. (1997). The Nature of External Representations in Problem Solving. *Cognitive Science, 21*(2), 179-217.

Zhang, J., & Norman, D. (1994). Representations in Distributed Cognitive Tasks. *Cognitive Science, 18*(1), 87-122.

## Appendix A: Examples from the questionnaires of Experiment 1 - 2

In the questionnaires we used, there were declarative knowledge questions concerning reproducing factual rule knowledge, and procedural knowledge questions concerning insight subjects had in the right strategy to solve the problem. There were also questions about opinions (statements that had to be scored on a 5-point Likert- scale).

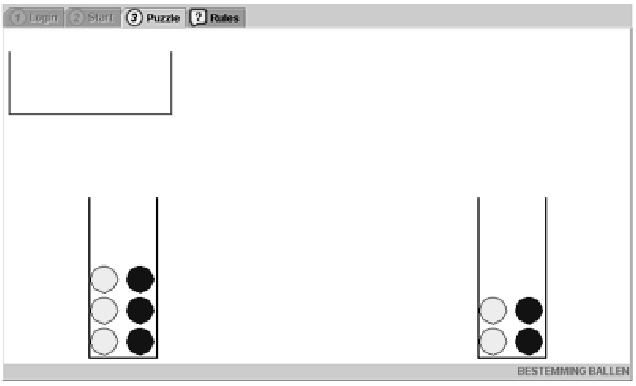
- **DECLARATIVE KNOWLEDGE QUESTIONS: AN EXAMPLE**

Welke van de volgende regels is juist: Na het voltooien van een zet,
Which of the following rules is correct: After doing a move,

O mogen er zich nooit meer blauwe dan gele ballen in een van de bakken bevinden.
*there can never be more blue balls than yellow balls in any of the two boxes*
O mogen er zich nooit meer gele dan blauwe ballen in een van de bakken bevinden.
*there can never be more yellow than blue balls in any of the two boxes*
O geen van bovenstaande regels is juist.
*none of the above is correct*

- **PROCEDURAL KNOWLEDGE QUESTIONS: AN EXAMPLE**

Bekijk het volgende screenshot.
*Have a look at the screenshot below*




Alle ballen moeten naar rechts worden gebracht. Wat is in deze spelsituatie de meest verstandige/logische vervolgtactiek?
*All balls have to be transported to the right side. What is the smartest / most logical move in this situation?*

O Alle blauwe ballen naar rechts
*All the blue balls to the right*
O Alle gele ballen naar rechts.
*All the yellow balls to the right*
O Evenveel gele als blauwe ballen naar rechts
*The same number of blue and yellow balls to the right*

- **OPINION QUESTIONS: A FEW EXAMPLES**

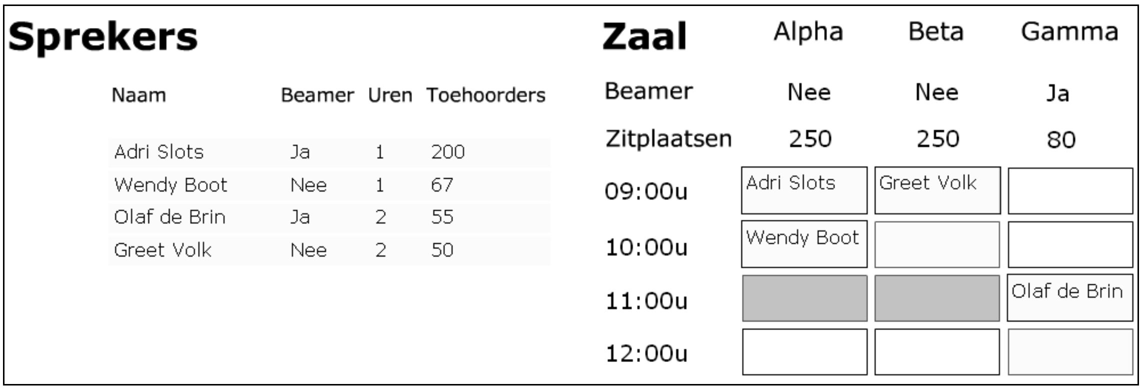
| | Geheel niet mee eens *Completely disagree* | Niet mee eens *Disagree* | Neutraal *Neutral* | Mee eens *Agree* | Geheel mee eens *Completely agree* |
|---|---|---|---|---|---|
| Het type probleem, dat in deze puzzel opgelost diende te worden, kwam mij bekend voor | O | O | O | O | O |
| Ik vond de taak moeilijk om te volbrengen *Performing the tasks was difficult* | O | O | O | O | O |
| Ik vond de taak plezierig om uit te voeren *The task was a pleasant one to perform* | O | O | O | O | O |
| De regels waren duidelijk genoeg om de taak te volbrengen *The rules were clear enough to perform the task* | O | O | O | O | O |
| Ik probeerde mijn stappen zoveel mogelijk van tevoren te plannen *I tried to plan my moves as much as possible before making a move* | O | O | O | O | O |

## Appendix B: Examples from the questionnaires of Experiment 3 - 5

In the questionnaires we used, there were declarative knowledge questions concerning judgments of situations, where rules of the problem are violated (or not). There were also procedural knowledge questions in which subjects were specifically asked about a procedure (strategy involved). There were also questions about opinions (statements that had to be scored on a 5-point Likert-scale).

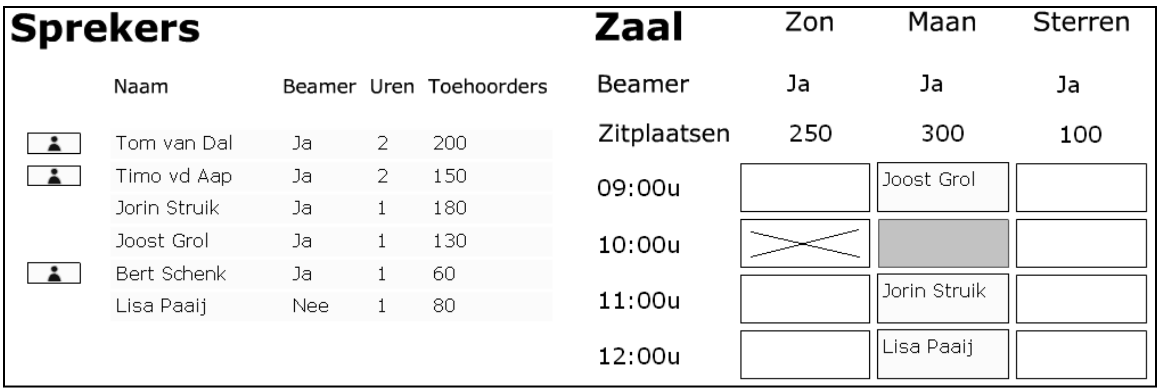
- **DECLARATIVE KNOWLEDGE QUESTIONS: AN EXAMPLE**

Bekijk het onderstaande rooster. Is dit een geldige planning? Waarom wel/niet?
*Look at the timetable below. Is this a valid schedule? Why (not)?*




- **PROCEDURAL KNOWLEDGE QUESTIONS: AN EXAMPLE**

Bekijk het onderstaande rooster. U wilt Timo vd Aap in de zaal Zon om 10:00u (aangegeven met een kruis) roosteren. Is dit een slimme zet? Geef ook aan waarom wel/niet.
*Look at the timetable below. You want to schedule Timo vd Aap in the Zon room at 10:00h (indicated with a cross). Is this a smart move? Also indicate why (not).*

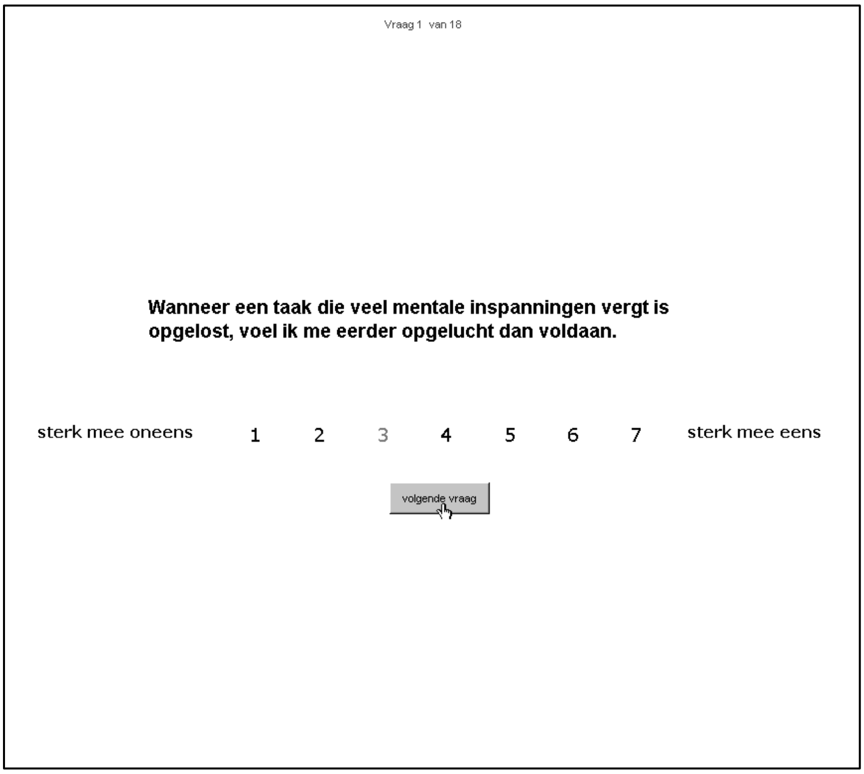


- **OPINION QUESTIONS: A FEW EXAMPLES**

| | Geheel niet mee eens<br><br>*Completely disagree* | Niet mee eens<br><br>*Disagree* | Neutraal<br><br>*Neutral* | Mee eens<br><br>*Agree* | Geheel mee eens<br><br>*Completely agree* |
|---|---|---|---|---|---|
| Ik wist steeds zeker of ik een bepaalde spreker in een bepaald vak kon plaatsen<br>*I always knew for sure whether I could place a certain speaker at a certain location.* | O | O | O | O | O |
| De bediening van dit programma werkt prettig<br>*The controls of this program are pleasant* | O | O | O | O | O |
| Ik had regelmatig het gevoel dat ik vastgelopen was en de oplossing niet kon vinden<br>*I often felt like I got stuck and could not find the solution* | O | O | O | O | O |
| De regels waren duidelijk genoeg om de taak te volbrengen<br>*The rules were clear enough to perform the task* | O | O | O | O | O |
| Ik vond de vijf taken lastig om op te lossen<br>*I found the five tasks difficult to solve* | O | O | O | O | O |

## Appendix C: The Need for Cognition Scale

The original 18 items from the short NFC scale of Cacioppo, Petty & Kao (1984) are listed below. Underneath each sentence is our Dutch translation. Items 3, 4, 5, 7, 8, 9, 12, 16 and 17 are reverse-scored. The questionnaire was administered with an application implemented with Macromedia Authorware (see screenshot).




1. I would prefer complex to simple problems.
   *Als ik mag kiezen, heb ik liever een complex dan een eenvoudig probleem.*

2. I like to have the responsibility of handling a situation that requires a lot of thinking.
   *Ik ben graag verantwoordelijk in een situatie waarin veel nagedacht moet worden.*

3. Thinking is not my idea of fun.
   *Denken is niet bepaald mijn idee van plezier hebben.*

4. I would rather do something that requires little thought than something that is sure to challenge my thinking abilities.
   *Ik doe liever iets waarbij weinig nagedacht hoeft te worden, dan iets waarbij mijn denkvermogen op de proef wordt gesteld.*

5. I try to anticipate and avoid situations where there is likely a chance I will have to think in depth about something.
   *Ik probeer situaties waarin ik waarschijnlijk diep moet nadenken te voorzien en te vermijden.*

6. I find satisfaction in deliberating hard and for long hours.
   *Iets langdurig en precies afwegen geeft me voldoening.*

7. I only think as hard as I have to.
   *Ik denk alleen zoveel als nodig is.*

8. I prefer to think about small, daily projects to long-term ones.
   *Ik denk liever na over kleine dagelijkse projecten dan over lange termijn projecten.*

9. I like tasks that require little thought once I've learned them.
   *Ik hou van taken waarbij weinig nagedacht hoeft te worden wanneer ik ze eenmaal geleerd heb.*

10. The idea of relying on thought to make my way to the top appeals to me.
    *Ik vind het een aantrekkelijk idee om op mijn verstand te vertrouwen bij het bereiken van mijn doel.*

11. I really enjoy a task that involves coming up with new solutions to problems.
    *Ik geniet echt van een taak waarin men met nieuwe oplossingen voor problemen moet komen.*

12. Learning new ways to think doesn't excite me very much.
    *Het leren van nieuwe manieren om te denken spreekt me niet bijzonder aan.*

13. I prefer my life to be filled with puzzles that I must solve.
    *Ik vind het leuk wanneer mijn leven gevuld is met puzzels die ik moet oplossen.*

14. The notion of thinking abstractly is appealing to me.
    *Het idee om abstract te denken vind ik aantrekkelijk.*

15. I would prefer a task that is intellectual, difficult, and important to one that is somewhat important but does not require much thought.
    *Ik heb liever een intellectuele, moeilijke en belangrijke taak, dan een die enigszins belangrijk is maar niet te veel denkwerk vereist.*

16. I feel relief rather than satisfaction after completing a task that required a lot of mental effort.
    *Wanneer een taak die veel mentale inspanningen vergt is opgelost, voel ik me eerder opgelucht dan voldaan.*

17. It's enough for me that something gets the job done; I don't care how or why it works.
    *Ik vind het voldoende wanneer iets blijkt te werken; hoe of waarom het precies werkt interesseert me niet.*

18. I usually end up deliberating about issues even when they do not affect me personally.
    *Gewoonlijk maak ik uitgebreid afwegingen over zaken, zelfs als die niet op mij persoonlijk betrekking hebben.*

## Appendix D: The strategy measure

In our tasks, speakers varied on a combination of variables: needed size of the room, needing a projector or not, or needing a 1 or 2 hour slot. One can of course make a typology of speakers by combining these variables, but a less complex way to look at it is to assign labels to speakers reflecting their "difficulty" in a simpler manner, corresponding with the nature of the grid where they have to be scheduled. There are always three rooms to schedule the speakers with their different constraints. Some speakers fit in all three the rooms, some in two of them, and the most difficult speakers fit only in one. Our measure "difficulty of speaker" has the values "1" (fits in only 1 room), "2" (fits in 2 rooms) and "3" (fits in all 3 rooms) where 1 is the most difficult type of speaker to schedule (lots of constraints) and 3 are the easiest speakers to schedule (no constraints, they fit anywhere). In the ideal case, first the speakers type 1, then type 2, and finally type 3 are scheduled. From the log files of the subjects, we can abstract of each task the sequence in which speakers with varying difficulty were placed. Below is an example of 2 scheduling sequences.

In these two examples there are 14 speakers:

- Type 1 speakers: 3
- Type 2 speakers: 6
- Type 3 speakers: 5

| **Example A** | **Example B** |
|---|---|
| 1 | 1 |
| 1 | 1 |
| 1 | 1 |
| 2 | 2 |
| 2 | 2 |
| 2 | 2 |
| 2 | 2 |
| 2 | 3 |
| 2 | 3 |
| 3 | 2 |
| 3 | 2 |
| 3 | 3 |
| 3 | 3 |
| 3 | 3 |

In the ideal case, when a perfect 1-2-3 sequence takes place, the strategy score is 100%, meaning that first all 1's, then all 2's and finally all 3's are scheduled (as in example A). This might be obvious, but let us demonstrate how this can be calculated. One can obtain a measure of this by answering two questions about each type of speaker (1, 2, or 3).

**Ideally, speakers of TYPE 1 are scheduled *before* all the speakers of type 2 & 3.**

1. Before how many speakers of type 2 was each type 1 speaker scheduled? (ideally before all: 6)
2. Before how many speakers of type 3 was each type 1 speaker scheduled? (ideally before all: 5)

**Ideally, speakers of TYPE 2 are scheduled *after* speakers of type 1 and *before* speakers of type 3.**

3. Before how many speakers of type 1 was each type 2 speaker scheduled? (ideally before none: 0)
4. Before how many speakers of type 3 was each type 2 speaker scheduled? (ideally before all: 5)

**Ideally, speakers of TYPE 3 are scheduled *after* all the other speakers (of type 2 & 3)**.

5. Before how many speakers of type 1 was each type 3 speaker scheduled? (ideally before none: 0)
6. Before how many speakers of type 2 was each type 3 speaker scheduled? (ideally before none: 0)

**Calculation for example A**

In our example A, the scores per speaker and the rest of the calculations come down to the following:

| | **Score Q1** | **Score Q2** |
|---|---|---|
| Speaker **TYPE1** | 6 | 5 |
| Speaker **TYPE1** | 6 | 5 |
| Speaker **TYPE1** | 6 | 5 |
| Total | **18** | **15** |

| | **Score Q3** | **ScoreQ4** |
|---|---|---|
| Speaker **TYPE2** | 0 | 5 |
| Speaker **TYPE2** | 0 | 5 |
| Speaker **TYPE2** | 0 | 5 |
| Speaker **TYPE2** | 0 | 5 |
| Speaker **TYPE2** | 0 | 5 |
| Speaker **TYPE2** | 0 | 5 |
| Total | **0** | **30** |

| | **Score Q5** | **ScoreQ6** |
|---|---|---|
| Speaker **TYPE3** | 0 | 0 |
| Speaker **TYPE3** | 0 | 0 |
| Speaker **TYPE3** | 0 | 0 |
| Speaker **TYPE3** | 0 | 0 |
| Speaker **TYPE3** | 0 | 0 |
| Total | **0** | **0** |

Then, these two scores per speaker type have to be summed in order to obtain 6 values, corresponding with the six questions above (see totals on the left). In Example A, this would mean:

| | |
|---|---|
| Q1: Speaker type 1 before type 2: | **18** |
| Q2: Speaker type 1 before type 3: | **15** |
| Q3: Speaker type 2 before type 1: | **0** |
| Q4: Speaker type 2 before type 3: | **30** |
| Q5: Speaker type 3 before type 1: | **0** |
| Q6: Speaker type 3 before type 2: | **0** |

Finally, These 6 values have to be entered in the matrix below. The percentage is calculated by dividing the total on left below the diagonal by the total for all the cells. In this case, the total left below the diagonal is 63, and the total for the complete matrix is also 63. 63/63 = 1 **(100%).**

| | Type 1 | Type 2 | Type 3 |
|---|---|---|---|
| ↓ | placed before | | |
| Type 1 | | 0 | 0 |
| Type 2 | 18 | | 0 |
| Type 3 | 15 | 30 | |

**Calculation for example B**

In our example B, the scores per speaker and the rest of the calculations come down to the following:

| | **Score Q1** | **Score Q2** |
|---|---|---|
| Speaker **TYPE1** | 6 | 5 |
| Speaker **TYPE1** | 6 | 5 |
| Speaker **TYPE1** | 6 | 5 |
| Total | **18** | **15** |

| | **Score Q3** | **ScoreQ4** |
|---|---|---|
| Speaker **TYPE2** | 0 | 5 |
| Speaker **TYPE2** | 0 | 5 |
| Speaker **TYPE2** | 0 | 5 |
| Speaker **TYPE2** | 0 | 5 |
| Speaker **TYPE2** | 0 | 3 |
| Speaker **TYPE2** | 0 | 3 |
| Total | **0** | **26** |

| | **Score Q5** | **ScoreQ6** |
|---|---|---|
| Speaker **TYPE3** | 0 | 2 |
| Speaker **TYPE3** | 0 | 2 |
| Speaker **TYPE3** | 0 | 0 |
| Speaker **TYPE3** | 0 | 0 |
| Speaker **TYPE3** | 0 | 0 |
| Total | **0** | **4** |

Then, these two scores per speaker type have to be summed in order to obtain 6 values, corresponding with the six questions above (see totals on the left). In Example A, this would mean:

| | |
|---|---|
| Q1: Speaker type 1 before type 2: | **18** |
| Q2: Speaker type 1 before type 3: | **15** |
| Q3: Speaker type 2 before type 1: | **0** |
| Q4: Speaker type 2 before type 3: | **26** |
| Q5: Speaker type 3 before type 1: | **0** |
| Q6: Speaker type 3 before type 2: | **4** |

When enter these 6 values have into the matrix below, the sum is different. The total on left below the diagonal is now 59, and the total for the complete matrix is still 63. The strategy score as percentage, is now 59/63 = 0.9365 **(93,65 %).**

| | Type 1 | Type 2 | Type 3 |
|---|---|---|---|
| | placed before | | |
| Type 1 | | 0 | 0 |
| Type 2 | 18 | | 4 |
| Type 3 | 15 | 26 | |

# Summary


## Background and research questions

Computer systems have seen a tremendous development, becoming more advanced and demanding fast learning from users. They are (often) friendly systems in which notions from Usability are implemented, by designing for effectiveness, efficiency, and satisfaction. Features such as visual feedback, guidance, and allowing for errors (undoing actions is possible) are often included. However, there are situations where excluding the chance of errors in the first place is important, instead of error recovery. In safety critical systems, safeguarding error prevention is as critical as users having the necessary knowledge at their immediate command. It can be desired that users are continuously proactive and do not operate in a shallow manner, caused by a system that gives them the feeling the task is being carried out for them. A fundamental question is whether it can be good to sometimes make a system that is *not* as assisting and guiding as one is used to.

There is a distinction between *plan-based* and *display-based* problem solving. Learning and using detailed plans for problem solving characterize the first type. *Display-based* problem solving uses interface information to structure and guide problem solving. In modern interfaces, it is common to work in a *display-based* manner and to rely on information the interface provides, where recognition rather than recall of commands drives interaction. A recurring issue in usability is the importance of minimizing user memory load to relieve the Working Memory of a user so that a maximum of cognitive resources can be devoted to the main task. To achieve this, information can be *externalized* and thus be of assistance. By *Externalization,* the necessity to remember characteristics of operators (e.g. limits on their applicability, consequences of applying them) is eliminated by displaying relevant information visually. When this is not the case, a user has to *internalize* and store relevant information, and this information is self-derived.

Research by Zhang and Norman (1994) showed advantages of external representations. They can provide constraints on the range of actions, which reduces the size of the problem space. It showed that externalized information could be beneficial in serving the user, resulting in better performance because it made tasks easier. However, there are also findings stating that making an interface harder (on the operator level) results in better performance (O'Hara & Payne, 1998; 1999). It appeared that when the cost associated with an operator was high, problem solving strategies became more *plan-based*, whereby search paths were evaluated mentally. When the cost associated with operators was lower, subjects shifted to a less planful strategy in which search was more reactive and *display-based*. If one thinks in general terms about the opposition between "making it easier" (alter representation for better performance) or "making it harder" (alter operator/controls for better performance) it could seem that the findings of the mentioned authors are contradicting. However, if one scrutinizes, it becomes clear that the two currents of research (alter representation vs. alter operator/controls) are not contradicting each other on the same dimension. We connected these currents and positioned our angle on the matter as being *between* the two. What is varied is how operators (parts of the interface: secondary task) become part of the primary task (solving the problem) by *dynamically* displaying relevant information. In our research, we did not vary the representation or the operators per se, but a combination of these two; we varied the external representation *via the operator*, the controls in the interface become the conveyors of externalized information, which tells a

user what objects can do *at this moment*, it externalizes the enforcement of rules. We used *destination feedback* as a means of Externalization. Destination feedback can guide decisions in showing which actions are possible, is quite common in modern computer interfaces, and has been advocated by authorities in providing guidelines (Sun, 2001; Apple, 1992).

This project investigates the conditions under which externalizing interface information by interface controls influences users' performance in solving problems requiring planning. In systems design, when user goals are such that not knowing or finding the right command as fast and directly as possible has severe consequences, one might consider more than plain usability. The notion that a friendlier system can result in more shallow behavior is worthwhile to delve into, for example in situations where knowledge construction *itself* or transfer of skill is important. One should take in account that not all interactive systems have *only* the goal of "ease of use from the first time on" and "maximum user satisfaction". We looked on a deeper level at what these presumed friendly systems do to humans and their behavior. Our *main research question* was: in tasks where planning is required, which interface style leads to more plan-based behavior, better strategy, and consequently better task performance? And besides immediate performance, which interface style causes better knowledge of the tasks and solutions afterwards in a transfer situation (with altered task/interface circumstances), or when a severe task interruption occurs?

We address these questions on *three levels*: firstly, a *behavioral* level concerning how interface style (Internalization or Externalization) influences user behavior and performance when learning to do a task, performing the task afterwards or under altered circumstances, and long-term memory of the task. Secondly there is a *theoretical* level concerning obtaining more insight in how, and to which extent plan-based vs. display-based problem strategies are provoked by the interface style. We try to gain insight in mental processes involved and how we can connect these to existing theoretical notions. Thirdly, there is a *practical* level involving how insights obtained by answers to the above can contribute to a predictive theory about when to apply Externalization features (or not) in interfaces. The objective is to propose an interaction framework with a trade-off between common usability on one side (designing systems that are pleasant to use, externalize information to aid users) and designing systems such that they keep users proactive (forcing them to internalize, avoid lazy behavior) on the other side. Generally, our hypothesis states that the Externalization of information leads to shallower processing and less proactive contemplation. Hence, we expect Externalization to result in less efficient performance on problem solving tasks. An interface where users have to internalize information without guidance, might seem like a more unfriendly system, but is expected to result in more efficient solutions and better imprinting of knowledge in memory.

Regarding the theoretical background, Human Computer Interaction research often resorts to notions from *Information Processing Psychology* (IPP). IPP describes various types of cognitive processes of human thought with the central claim that humans operate as an information processing system, and cognition is largely regarded as computation. However, this conception might be too rational and mechanical, assuming obedient problems solvers in isolated situations, and as such too limited for our HCI context. Since our focus is on how to provoke users into certain kinds of cognitive activity, we adopted *Cognitive Load Theory* (CLT) next to the information processing account (Sweller, 1988).

CLT provides guidelines to assist in the presentation of information in a manner that encourages optimized learner activities and performance and is concerned with limitations of Working Memory. CLT has the basic premise that learning is hindered if the instructional material overwhelms learners' cognitive resources. CLT distinguishes Intrinsic, Germane, and Extraneous Cognitive load. Intrinsic Cognitive Load is imposed by the task and refers to the complexity of a concept itself. Extraneous Cognitive Load results from the design of the material, and refers to *unnecessary* load caused by inefficient design. Germane Cognitive Load refers to the degree of effort involved in processes that are effective for learning, and is associated with motivation and interest. Learning is optimized when a good portion of Working Memory can be devoted to effortful cognitive processes. We propose that Externalization (by destination feedback) takes away resources from the task; thereby increasing Extraneous Cognitive Load causing Germane Cognitive Load to be low, because there is less bandwidth left in Working Memory. Furthermore, the suggestion of assistance makes users less willing to invest effort and make inferences. In contrast, Internalization is thought to cause low Extraneous Cognitive Load, since no resources are taken away from the task and processing interface events such as feedback does not distract users. We associated having to *internalize* information with Germane Cognitive Load, directed at effortful and beneficial cognitive processes.

**Results and conclusions on the behavioral level**

By adopting conceptions of the CLT framework and connecting these to the concepts of plan-based and display-based behavior and how this is influenced by interface styles where Internalization and Externalization is varied, we attempt to address our research questions on the mentioned three levels.

To answer our questions on the *behavioral* level concerning task performance, two series of controlled experiments (5 in total) were conducted using two interface styles: one version in which certain information is externalized onto the interface (Externalization) and another version where this is not done (Internalization). In the Externalization version, the operators in the interface conveyed information, in the Internalization version this was not the case. The first series of experiments used a computerized isomorphic version of the well-known "Missionaries & Cannibals" problem. The second series of experiments used a more realistic office-like application called "Conference Planner". Several measures were collected, among which the solution path (superfluous moves, an important measure reflecting planning and strategy), timestamps and correctness. We also assessed user knowledge and studied eye tracking data.

In the first series of two experiments, we varied interface styles in a puzzle task that requires planning called "Balls & Boxes", an isomorphic version of the Missionaries & Cannibals problem. Users solved problems in two conditions: an Externalization interface style, which provided information about the legality of actions, or an Internalization interface style where this was absent. We investigated which interface style resulted in better performance, knowledge about the tasks and solutions afterwards, and the long-term effects after eight months. In addition, we compared performance in a transfer situation, and lastly whether the influence of instructing subjects to plan had different effects in the two interface styles.

The results showed that subjects who worked with the Internalization version solved more puzzles. Overall, the time needed for the task was practically the same in both interface versions. After a break, Internalization subjects performed less superfluous moves and had better knowledge afterwards. In a second session with the same subjects, after a large delay, Internalization subjects had a better memory of same task, and they performed better on a transfer task. Another finding was that when a planning instruction was presented, only the Internalization subjects followed it. We conclude that Internalization subjects invested more effort in actively trying to come up with a solution (plan-based behavior), which is in contrast with more display-based behavior, which we presume the Externalization subjects applied. Also, Internalization subjects indeed imprinted strategies better, since they had solutions more ready available after a delay, and were better able to apply their knowledge in a transfer situation.

For the second series of three experiments, we developed an application (Conference Planner) that was more like realistic computer tasks in daily life. It was a constraint-satisfaction scheduling task, involving planning speakers at a conference. Also of this task, an Externalization interface and an Internalization interface were constructed. Externalization was implemented by highlighting legal options in the Externalization condition where a speaker can be placed. Note that this did not show the best slot to place a speaker, but simply which slots are possible. In addition, a transfer version was developed, called Ferry Planner, which had the same underlying task structure, but felt and looked different (the task involved planning coaches on a ferry deck). In the three experiments, we investigated as before how performance was influenced by interface style. We also checked whether the cognitive style already present in subjects measured by the Need for Cognition scale influenced performance in the two interface versions. In addition, we investigated transfer of skill by letting subjects switch to another task (Ferry Planner), to another interface style (the opposite of which they used before) or switch both task and interface style at the same time. The aim was to see whether the interface style one worked with before influenced performance on another task. Lastly, we looked at the effect of a severe interruption. We assumed that when actions are more planned (as we presume is the case in the Internalization version), task resumption is smoother, and more error free than in the Externalization version.

The second series of experiments showed again that Internalization subjects were at advantage. There were no *overall* time differences. However, Internalization subjects took more time before they started solving the problems, and between the individual moves. This is thought to reflect planning and contemplation. Furthermore, Internalization subjects performed less superfluous moves. Preexisting cognitive style of users had no influence on behavior and performance; it seemed that only our manipulation caused the differences. We also introduced an interruption that was severe and highly disrupting (a system crash). The Internalization subjects were better at task resumption after this severe interruption. Internalization subjects continued as if nothing happened, whereas the performance of the Externalization subjects deteriorated badly. We concluded from this that Internalization subjects have worked in a more plan-based manner and invested more effort in forming strategies. This caused better resistance to a severe interruption. Regarding transfer of skill, switching *interface* in the sequence INT→EXT always resulted in better performance than EXT→INT, and switching to *another (transfer) task* in the sequence EXT→INT caused worse performance than other combinations. However, the effect of interface subjects worked with did not last for long, it disappeared after the second transfer task. The eye tracking data showed some differences in gaze patterns between the two conditions, but not everywhere where we expected it. In the first tasks, Externalization subjects spent more time looking at the area where externalized feedback occurred, and where the important information was *not* located. We also formally analyzed the strategy that subjects applied. Internalization subjects applied a most-constraints-first approach more, which is the smartest way to an economical solution.

Resuming the results of all the experiments on the behavioral and performance level, it showed that Externalization as we implemented it did not have any beneficial effect. Immediate and delayed performance was worse with the Externalization interface. In addition, transfer of skill was worse for these users, both to another task, and to another interface. Having to internalize information oneself, and not relying on information from the interface caused subjects to be more proactive and plan their actions more. Internalization subjects imprinted relevant task and rule knowledge better and were not affected by a severe interruption in the workflow, whereas Externalization subjects were.

**Conclusions on the theoretical level**

On a *theoretical* level, we obtained more insight in how plan-based vs. display-based problem behavior is provoked by the Internalization respectively Externalization interface style. We conclude that our Internalization interface causes only low Extraneous Cognitive Load, since no resources are taken away from the task and users are not distracted by interface events, no Working Memory resources have to be devoted to them. Also a cost-benefit consideration plays a part; Internalization subjects realize quickly that aimlessly clicking around will not help. Plan-based behavior is provoked, triggering users to invest more focused effort. This corresponds with high Germane Cognitive Load (directed at beneficial processes and associated with motivation). Internalization subjects study the situation better, are less distracted, and follow a more solid plan. The result of Internalization subjects making more inferences and their smarter strategies is less trial and error problem solving, leading to more straightforward solutions (efficiency). Doing the mental processing more by oneself causes better imprinted knowledge and insight immediately, and also after a large delay. Because users have internalized information and constructed strategies oneself, knowledge is also better applicable to similar problems in transfer situations. The Externalization Interface might allow cognitive offloading, but this in itself does not cause users to devote more effort to beneficial mental processes. The destination feedback provided users with the feeling that they were guided, and that the task is partially carried out for them. This discourages active contemplation, it makes users less willing to invest effort and make inferences. They behave lazier and shallower, Germane Cognitive Load is low, reflection is not encouraged, and more display-based behavior occurs. This also overrides the personal cognitive style of users. Besides this, the destination feedback still has to be processed by the user, and can be regarded as Extraneous Cognitive Load. It brings nothing positive; it acts as a distracter, and inhibits the focus on the problem itself. When users depend to a high degree on the externalized feedback, behavior is largely display-based, accompanied by unnecessary actions that steer away from a direct solution.

Resuming, Externalization makes users count on the interface and gives them the feeling (unrightfully so) that the thinking-work is done for them. This seduces them into more shallow cognitive behavior and discourages undertaking cognitive activities aimed at strategy and knowledge construction. Users who internalize information themselves behave more plan-based, invest more effort in cognitive processes, and are more proactive and ready to make inferences. This in turn results in more focus, more direct and economical solutions, better strategies, and better imprinting of knowledge. This knowledge is easier to recall at a future point in time, and is better transferable to transfer situations where the interface, the task, or both are different, and less vulnerable to a severe interruption. Regarding the combination of theoretical concepts we mentioned, we concluded that CLT is valuable to HCI, but it cannot by itself account for the complete range of processes that take place, and neither can Information Processing Psychology. CLT as a theoretical concept is valuable for describing events and issues that happen in the periphery of problem solving in computer-based tasks. However, the actual informational processes that problem solvers engage in are incomplete without notions from IPP. Concepts from CLT, such as Extraneous and Germane Cognitive Load can be used to describe what happens with a user in certain circumstances, how interface-related events can change our behavior in approaching a task, especially concerning the motivation and effort involved. Alongside the notion of cost-benefit, we see Cognitive Load as influencing Information Processing processes.

**Conclusions on the practical level**

We conclude that there is more to HCI guidelines than how they are now. Computer-mediated tasks can take advantage of interfaces that are designed from considerations that run deeper than plain usability, even when they go against common sense. Our experiments concerned a specific kind of Externalization, which is very common: destination feedback. Relieving a user's memory and making interactions assisted by destination feedback had negative effects. It made users count on the interface and gave them (unrightfully so) the feeling that the task and thinking-work is partly done for them and seduced them into more shallow cognitive behavior. In designing interfaces we have to be careful with providing interface cues that give away too much, and must design in such a manner that the way users (should) think is optimally supported, which in turn could help the software to achieve its specific goal. We presented an interaction framework, which can guide decisions *outside* the category of tasks where plain common usability is high on the agenda, without presuming our framework it is complete regarding all task environments. It can guide decisions regarding interface design and interactive interface properties in more serious contexts, such as educational contexts where learning itself is the aim or safety critical situations. It also can be important when dependence on a particular interface is specifically undesired, such as when risky and complex tasks are performed, or when a user suddenly is confronted with a new situation and his knowledge has to be applied in a different situation. The common guideline to "not give users the chance to make mistakes" should of course not be neglected, but at the same time, interaction should facilitate or even persuade users to learn what underlies the task. The same is true in situations where interruptions are commonplace and where mastery of what is underlying a task or domain is desired, or when operations come with a cost and direct solutions are the aim.

**Overall conclusions**

Immediate and delayed performance when using our Externalization interface, which had destination feedback, was worse than performance using the Internalization interface. Also transfer of skill was worse for users of the Externalization interface, both to another task, and to another interface. These users were characterized by display-based behavior. Subjects that used the Internalization interface imprinted relevant task and rule knowledge better and were not affected by a severe interruption in the workflow, whereas Externalization subjects were. We conclude that users who internalize information themselves behave more plan-based, and are more proactive and ready to make inferences. This in turn results in more focus, more direct and economical solutions, better strategies, and better imprinting of knowledge.

This knowledge is easier to recall at a future point in time, less vulnerable to a severe interruption in the workflow, and better transferable to transfer situations where the interface, the task, or both are different.

Human-computer interaction designers can take advantage from these considerations that go beyond plain usability, even when they go against common sense. In designing interfaces we have to take care with providing interface cues that give away too much information, and must design in such a manner that the way users (should) think is optimally supported, which in turn could help the software to achieve its specific goal. Examples are situations where risky and complex tasks are performed, or where a user suddenly is confronted with a new situation. One can also think of situations in which interruptions are commonplace, or where operations come with a cost and direct solutions without deviations are the aim. Based on our findings an interaction framework is proposed that can guide decisions regarding interface design.

# Samenvatting

## Achtergrond en onderzoeksvragen

Computer soft- en hardware hebben de laatste decennia een enorme ontwikkeling doorgemaakt, worden steeds geavanceerder, en eisen dat gebruikers er snel mee leren om te gaan. Het zijn vaak zijn systemen die prettig in gebruik zijn, waarin de noties van "gebruiksvriendelijkheid" zijn geïmplementeerd. Eigenschappen die vaak ingebouwd worden, zijn visuele feedback, assistentie, en het toestaan van fouten, in de zin dat ze ongedaan kunnen worden gemaakt. Er zijn echter ook situaties denkbaar waarin niet zozeer het ongedaan maken, of het herstellen van fouten belangrijk is, maar waar het in de eerste plaats zaak is het maken van fouten gewoonweg te voorkomen. Men kan denken aan systemen waar veiligheid een rol speelt en waar foutloos gebruik cruciaal is. Het kan gewenst zijn dat gebruikers zich steeds proactief gedragen en niet op een oppervlakkige manier interacteren met het systeem, doordat het systeem de gebruiker de indruk geeft dat het werk (deels) gedaan wordt voor de gebruiker. Een centrale vraag is dan of het soms niet goed kan zijn een systeem zo te maken dat het minder assisteert dan we gewend zijn.

Er wordt onderscheid gemaakt tussen *plangebaseerd* en *schermgebaseerd* probleem oplossen. Het eerste type kenmerkt zich door het aanleren en gebruiken van gedetailleerde plannen. Bij schermgebaseerd probleem oplossen wordt er juist informatie uit de computer interface gebruikt om het proces te sturen. In hedendaagse interfaces is het gebruikelijk om op een schermgebaseerde wijze te werk te gaan en te vertrouwen op de informatie die de interface geeft. In deze situatie is het herkennen, en niet zozeer het herinneren van commando's dat de interactie stuurt. Een terugkerend onderwerp in onderzoek naar gebruiksvriendelijkheid is het laag houden van geheugenbelasting met als doel het werkgeheugen van een gebruiker te ontzien zodat de beschikbare cognitieve capaciteit maximaal benut kan worden voor de hoofdtaak. Om dit te bereiken, kan informatie worden *geëxternaliseerd,* en zo de gebruiker assisteren. Door *Externalisatie* is de noodzaak om zich de specifieke eigenschappen van bedieningsoperatoren te herinneren (bij voorbeeld de grenzen van hun toepasbaarheid, de consequenties ervan) weggehaald doordat relevante taakinformatie getoond wordt. Wanneer dit niet gedaan wordt, moet een gebruiker *zelf* relevante informatie *internaliseren* en opslaan in zijn geheugen.

Onderzoek van Zhang en Norman (1994) toonde aan dat het extern representeren van informatie voordelen kan hebben. Het kan het aantal opties voor beperken, waardoor de probleemruimte kleiner wordt. Het bleek dat geëxternaliseerde informatie een positief effect had omdat het de gebruiker hielp en de taken gemakkelijker maakte, wat resulteerde in betere prestaties. Er zijn echter ook onderzoekers die beweren dat als een interface juist moeilijker wordt gemaakt op het niveau van de bedieningsoperatoren, dit tot betere prestaties leidt (O'Hara & Payne, 1998; 1999). Zij toonden aan dat wanneer de kosten die aan een bedieningsoperator vastzaten hoog waren, de oplosstrategieën meer plangebaseerd werden, en gekarakteriseerd werden door mentale evaluatie van zoekpaden. Als de kosten van bedieningsoperatoren lager waren, gingen de gebruikers over op een minder planmatige, meer schermgebaseerde aanpak.

Als men in het algemeen denkt aan de tegenstelling tussen "makkelijker maken" (varieer de representatie voor betere prestatie) en "moeilijker maken" (varieer bedieningsoperatoren voor betere prestatie) dan lijken de bevindingen van de genoemde auteurs elkaar tegen te spreken. Als men dit echter van dichterbij bekijkt, wordt het duidelijk dat deze twee onderzoeksstromingen (variëren representatie vs. variëren bedieningsoperatoren) elkaar niet tegenspreken op dezelfde dimensie. Wij hebben deze twee zienswijzen gecombineerd, en positioneren ons tussen deze twee in. Wat we variëren in dit onderzoek, is hoe bedieningsoperatoren (delen van de interface die de secundaire taak vormen) onderdeel worden van de primaire taak (het oplossen van het probleem) door op dynamische wijze relevante informatie te tonen aan de gebruiker. Wij variëren echter niet de externe representatie van de bedieningsoperatoren zelf, maar variëren een combinatie; we variëren de externe representatie *via* de bedieningsoperatoren, in de zin dat de bedieningsoperatoren zelf de doorgever van informatie worden en zo de gebruiker vertellen wat de mogelijkheden zijn van een functie op dit moment. Het externaliseert de toepassing van de onderliggende regels. We gebruiken *destination feedback* als een manier om te externaliseren. Destination feedback kan als leidraad dienen tijdens het maken van beslissingen doordat het duidelijk maakt welke acties op dat moment ondernomen kunnen worden, en dit is eigenlijk een standaard techniek geworden in moderne computer interfaces, en het implementeren ervan wordt gepropageerd door grote marktpartijen die bij het opstellen van richtlijnen betrokken zijn (Sun, 2001; Apple, 1992).

In dit project onderzoeken we de voorwaarden waaronder het externaliseren van informatie via de bedieningsoperatoren het gedrag van de gebruiker kan beïnvloeden tijdens het probleem oplosproces bij taken waarin plannen belangrijk is. Als in het ontwerpen van systemen het gebruikersdoel zodanig is dat het zich niet herinneren of het niet snel kunnen vinden van het juiste bedieningscommando serieuze consequenties heeft, dan zou men overwegingen kunnen maken die verder reiken dan “gewone” gebruiksvriendelijkheid. Het idee dat een vriendelijker systeem kan resulteren in oppervlakkiger gedrag is de moeite van het onderzoeken waard, daar lang niet alle interactieve systemen “gebruiksgemak” vanaf het eerste moment dat men begint en “maximale tevredenheid” van de gebruiker als doel hebben. Wij hebben op een dieper niveau gekeken naar wat deze verondersteIde vriendelijke systemen teweegbrengen bij mensen en wat het doet met hun gedrag tijdens het gebruik. Onze *belangrijkste onderzoeksvraag* was: in taken waar planning noodzakelijk is, welke interface stijl lokt meer plangebaseerd gedrag uit en leidt tot een betere strategie met als gevolg een betere prestatie? En naast de prestatie op de korte termijn, welke interface stijl leidt tot betere kennis over de taak en betere toepasbaarheid van die kennis op een vergelijkbaar probleem (transfer of skill), of wanneer er een interruptie plaatsvindt?

We beantwoorden deze vragen op *drie niveaus*: eerst op het *gedrags* niveau dat ingaat op de invloed van interface stijl (Internalisatie versus Externalisatie) op gedrag en prestatie tijdens het leren van de taak, het doen van diezelfde taak op een later moment of onder andere omstandigheden, en het lange termijn geheugen voor de taak. Als tweede is er het *theoretisch* niveau waarin het doel is inzicht te krijgen in hoe, en in welke mate plangebaseerde versus schermgebaseerde probleem oplosstrategieën door de respectievelijke interfaces worden veroorzaakt. We proberen inzicht te krijgen in de mentale processen die hierin een rol spelen, en hoe we deze kunnen verbinden aan reeds bestaande theoretische inzichten. In de derde plaats is er een *praktisch* niveau betreffende de wijze waarop inzichten opgedaan door eerdergenoemde vragen te beantwoorden, kunnen bijdragen aan een beslissingen over wanneer wel, of juist niet Externalisatie eigenschappen te implementeren. Het doel is om een interactie-raamwerk te introduceren met daarin de afweging tussen gebruiksvriendelijkheid aan de ene kant (d.m.v. Externalisatie systemen zodanig ontwerpen die prettig zijn in het gebruik) en aan de andere kant het zodanig ontwerpen dat de gebruikers proactief blijven (ze prikkelen om zelf te internaliseren, en voorkomen van oppervlakkig gedrag). Onze algemene hypothese stelt dat het externaliseren van informatie leidt tot het oppervlakkiger verwerken van informatie, en minder proactieve overwegingen van de gebruiker. Hierdoor verwachten we dat Externalisatie resulteert in minder efficiënte prestatie op probleem oplostaken. Een interface waarbij de gebruikers zelf informatie moeten internaliseren moge onvriendelijker lijken, de verwachting is dat het leidt tot meer efficiënte oplossingen en betere opslag van de taakkennis in het geheugen.

Wat betreft de theoretische achtergrond, onderzoek op het gebied van Human Computer Interaction (HCI: Mens Machine Interactie) grijpt vaak terug op begrippen uit de *Informatie Verwerkings Psychologie (IVP)*. IVP beschrijft cognitieve processen in het menselijk denken met de centrale aanname dat mensen te werk gaan als een informatieverwerkend systeem, en cognitieve processen worden veelal beschouwd als computationele processen. Deze gedachte is echter misschien te rationeel en mechanisch, en gaat te veel uit van “gehoorzame” gebruikers die in geïsoleerde situaties taken uitvoeren, en is daarom te gelimiteerd voor onze HCI context. Omdat onze aandacht uitgaat naar de wijze waarop we gebruikers tot specifieke soorten gedrag aan kunnen zetten, gebruiken we de *Cognitive Load Theory* (Cognitieve Belasting Theorie) van Sweller (1988). Cognitive Load Theory (CLT) voorziet in richtlijnen om te helpen dat informatie representatie zodanig is dat het leer activiteiten stimuleert, en houdt zich bezig met de beperkingen van het werkgeheugen. Een centrale aanname van CLT is dat leren gehinderd wordt als het leermateriaal de cognitieve bronnen van de persoon die leert te boven gaat. CLT maakt onderscheid tussen “Intrinsic” (taakcomplexiteit), “Germane” (effectief), en “Extraneous” (ineffectief) Cognitive Load. Intrinsic Cognitive Load wordt veroorzaakt door de complexiteit van een taak zelf. Extraneous Cognitive Load komt voort uit het ontwerp van het leermateriaal, en heeft betrekking op *onnodige* cognitieve belasting door inefficiënt ontwerp. Germane Cognitive Load is de mate van inspanning of moeite tijdens de taakuitvoering, en dit staat in verband met motivatie en interesse. Leren is optimaal wanneer een adequaat deel van het werkgeheugen kan worden aangewend voor inspanningsrijke cognitieve processen. We nemen aan dat Externalisatie de mentale capaciteit voor de hoofdtaak verlaagt (destination feedback moet hoedanook ook cognitief verwerkt worden), en daarmee de Extraneous Cognitive Load verhoogt omdat er minder bandbreedte over is in het werkgeheugen. Doordat er gesuggereerd wordt dat men geassisteerd wordt, zijn gebruikers bovendien in mindere mate bereid moeite te investeren. Hier tegenover staat Internalisatie, dat geacht wordt lage Extraneous Cognitive Load te veroorzaken, omdat er immers geen capaciteit is weggehaald bij de taak, en het mentaal verwerken van interface gebeurtenissen leidt de gebruiker niet af. We associëren het internaliseren van informatie dan ook met Germane Cognitive Load, dat gericht is op intensieve en positieve cognitieve processen.

**Resultaten en conclusies op gedragsniveau**

We proberen onze onderzoeksvragen te beantwoorden door de begrippen uit CLT te gebruiken en deze te koppelen aan de concepten plan- en schermgebaseerd gedrag (en hoe Externalisatie resp. Internalisatie van informatie dit beïnvloeden).

Op het gedragsniveau zijn er twee reeksen van (in totaal vijf) experimenten uitgevoerd. Er werden steeds twee interface stijlen gebruikt. In de Externalisatie versie wordt bepaalde informatie geëxternaliseerd in de interface doordat de bedieningsoperatoren zelf informatie communiceerden. In de Internalisatie versie was dit niet het geval. In de eerste reeks experimenten werd een abstracte variant van het bekende “Missionarissen en Kannibalen” probleem gebruikt. In de tweede reeks experimenten werd een programma gebruikt dat meer overeenkwam met de manier waarop computerprogramma’s in het dagelijkse leven worden gebruikt (Conferentie Planner). Diverse zaken werden gemeten, waaronder het oplossingspad (het aantal overbodige zetten, een maat die relevant is voor plannen en strategie), tijdmetingen en correctheid. We hebben ook de kennis die men er van verkregen had in kaart gebracht, en bestudeerd waar gebruikers keken op het scherm.

In de eerste reeks (van twee) experimenten, varieerden we interface stijl in een puzzeltaak waarin plannen essentieel is (“Balls & Boxes”, een isomorfe versie van het “Missionarissen en Kannibalen“ probleem). Deze taak was er in twee condities: de Externalisatie interface waarin steeds te zien was of acties geoorloofd waren, en de Internalisatie interface stijl, waar deze informatie ontbrak. We onderzochten welke interface stijl resulteerde in betere prestaties op de taak, kennis over de taak en de oplossing, en de lange termijn effecten. Tevens hebben we de prestatie op een transfer taak vergeleken, en hebben we getest of het specifiek instrueren van deelnemers om goed na te denken verschillende effecten had bij de twee interface stijlen. Het bleek dat proefpersonen die met de Internalisatie versie gewerkt hadden meer puzzels oplosten. De tijd die men nodig had voor correct opgeloste puzzels was praktisch hetzelfde. Na een pauze maakten Internalisatie proefpersonen minder overbodige zetten, en deze personen hadden ook betere kennis achteraf. In een vervolgsessie met dezelfde proefpersonen die plaatsvond na acht maanden, hadden de Internalisatie proefpersonen de taak en de oplossing beter onthouden, en presteerden ze ook beter op een transfer taak. Verder bleek dat wanneer de specifieke plan-instructie gegeven werd, alleen de Internalisatie personen deze opvolgden. We concluderen dat de Internalisatie proefpersonen meer tijd en moeite investeerden in het op actieve wijze aandragen van een oplossing (plangebaseerd gedrag), wat het tegenovergestelde is van schermgebaseerd gedrag, waarvan we veronderstellen dat de Externalisatie proefpersonen erdoor gekenmerkt werden. De oplosstrategieën werden ook beter opgeslagen door Internalisatie personen, ze hadden de oplossing beter beschikbaar na een lange onderbreking, en ze waren beter in staat deze kennis toe te passen in een transfer situatie.

Voor de tweede reeks (van drie) experimenten werd een programma ontwikkeld dat meer leek op een computer taak uit het dagelijks leven. Met het programma (Conference Planner) dienden sprekers ingepland te worden op een conferentie. Het programma was er in twee versies: een Internalisatie versie en een Externalisatie versie. Externalisatie was geïmplementeerd door steeds te laten zien wat de toegestane opties waren (waar een spreker geplaatst kon worden). Dit liet echter alleen zien welk vak in het rooster mogelijk was, niet welk vak het slimste was met oog op de verdere planning. Ook hebben we een transfer versie ontwikkeld die wat onderliggende taak en structuur betreft hetzelfde was maar anders oogde en aanvoelde. Met deze drie experimenten hebben we opnieuw onderzocht hoe de prestatie op de taak beïnvloed wordt door interface stijl. Hier hebben we ook onderzocht of de al aanwezige cognitieve stijl in proefpersonen, gemeten met de Need for Cognition test (behoefte aan nadenken en puzzelen) van invloed is op prestatie bij de twee interface stijlen. Ook hebben we gekeken naar transfer of skill naar een ander taak, naar een ander interface stijl, of naar beide tegelijk. Het doel was erachter te komen of de interface stijl waarmee men eerder werkte de prestatie op een andere taak beïnvloedt. Als laatste hebben we gekeken naar het effect van een drastische onderbreking tijdens de taak. We namen aan dat wanneer acties meer gepland zijn (zoals we veronderstelden bij Internalisatie), taakhervatting soepeler verloopt, en gepaard gaat met minder fouten.

De tweede reeks experimenten liet opnieuw zien dat de Internalisatie proefpersonen in het voordeel waren. Hoewel er ook hier geen significante algemene tijdsverschillen waren, namen de Internalisatie proefpersonen meer tijd om na te denken voordat ze begonnen aan de puzzel en ook tussen de afzonderlijke stappen. Dit wijst op het plannen van acties. Voorts maakten de Internalisatie proefpersonen minder overbodige zetten in het oplossingspad. De cognitieve stijl van proefpersonen had geen invloed op gedrag en prestatie. Het lijkt erop dat alleen de manipulatie het gedrag beïnvloedde. De drastische onderbreking tijdens de taak die de proefpersonen ondergingen bestond uit een computercrash. De Internalisatie proefpersonen waren beter in staat de taak te hervatten na deze computercrash. De prestatie van de Externalisatie proefpersonen was veel slechter dan voorheen. We concluderen dat de Internalisatie proefpersonen te werk gingen op een plangebaseerde manier. Ze investeerden meer moeite in het bepalen van een oplosstrategie, die zorgde voor betere bescherming tegen een onderbreking zoals deze computercrash. Wat transfer of skill betreft, omschakelen naar *een andere interface* in de volgorde INT→EXT resulteerde altijd in betere prestatie dan EXT→INT. Omschakelen naar *een andere taak*, in de volgorde EXT→INT zorgde voor slechtere prestatie dan andere combinaties. Dit effect van interface stijl duurde echter niet lang, na een tweede transfer taak verdween het. De oogbewegingen van de proefpersonen naar het scherm waren op enkele punten verschillend tussen de twee interface stijlen, maar niet overal waar we het verwachtten. In het begin van de taken, keken Externalisatie proefpersonen langer en vaker naar die regio's van het scherm waar de feedback zichtbaar was, en waar de belangrijke informatie om zelf de oplossing te bepalen zich juist *niet* bevond. De formele analyse van de strategieën wees uit dat de Internalisatie proefpersonen strategieën toepasten waarin conferentiesprekers die moeilijk te plaatsen waren eerst geplaatst werden. Dit type strategieën vormt de meest economische weg naar een directe, efficiënte oplossing.

Samenvattend, de resultaten van de experimenten op het gedragsniveau toonden aan dat Externalisatie zoals hier geïmplementeerd geen enkel positief effect had. Prestatie op het moment zelf, en na verloop van tijd, was beter als er met de Internalisatie interface was gewerkt. Tevens was voor deze proefpersonen de transfer of skill beter naar een andere taak, en naar een andere interface stijl. De noodzaak om informatie zelf te internaliseren, en niet aangewezen zijn op informatie die de interface geeft, zorgde ervoor dat gebruikers proactiever en meer plangebaseerd te werk gingen. Internalisatie proefpersonen sloegen de relevante taak- en regelkennis beter op, en waren niet benadeeld door een drastische onderbreking terwijl dit laatste voor Externalisatie proefpersonen wel het geval was.

**Conclusie op het theoretische niveau**

Op theoretisch niveau hebben we meer inzicht gekregen in hoe plangebaseerd versus schermgebaseerd gedrag gestimuleerd kan worden door de Internalisatie dan wel Externalisatie interface stijl. We concluderen dat onze Internalisatie Interface slechts lage Extraneous Cognitive Load veroorzaakt, doordat er immers geen capaciteit van het werkgeheugen weggehaald wordt bij de taak zelf, en de gebruiker niet afgeleid wordt door gebeurtenissen in de interface. Ook een "kosten–voordeel" afweging speelt een rol; Internalisatie proefpersonen realiseerden zich snel dat doelloos in het rond klikken niet tot een oplossing leidt. Men wordt geprikkeld het probleem plangebaseerd aan te pakken, wat meer gerichte inspanning stimuleert. Dit komt overeen met Germane Cognitive Load, dat gericht is op zinvolle positieve processen die ook met motivatie te maken hebben. Internalisatie proefpersonen bestuderen de situatie beter, zijn minder afgeleid, en volgen een plan dat meer solide is. Dit zorgt ervoor dat men meer gevolgtrekkingen zelf maakt, hun strategie is minder “trial and error”, wat tot directere oplossingen leidt. Het zelf uitvoeren van de mentale processen zorgt ervoor dat kennis beter wordt opgeslagen en leidt tot beter inzicht. Precies omdat gebruikers informatie geïnternaliseerd hebben, en zelf strategieën ontwikkeld hebben, is deze kennis beter toepasbaar in gelijksoortige problemen in transfer situaties. De Externalisatie interface mag dan misschien soms cognitieve belasting verminderen, dit alleen zorgt er niet voor dat de gebruiker zijn inspanning investeert in positieve mentale processen. De destination feedback gaf gebruikers het gevoel dat ze geleid werden en dat de taak gedeeltelijk voor hen gedaan werd. Dit ontmoedigt actief overdenken, het maakt gebruikers minder bereid tot inspanning en het maken van gevolgtrekkingen. Ze gedragen zich oppervlakkiger, Germane Cognitive Load is laag, reflectie wordt niet gestimuleerd, en meer schermgebaseerd gedrag vindt plaats. Dit speelt een belangrijkere rol dan de al aanwezige cognitieve stijl van personen. Verder geldt ook dat de destination feedback hoe dan ook verwerkt dient te worden, en kan ook gezien worden als Extraneous Cognitive Load. Het brengt niets positiefs; het leidt af, en hindert de focus op het op te lossen probleem. Als gebruikers in hoge mate leunen op de geëxternaliseerde feedback, vertoont men schermgebaseerd gedrag dat gepaard gaat met onnodige acties die afleiden van een directe oplossing.

Samenvattend, Externalisatie zorgt ervoor dat gebruikers leunen op de interface, en geeft hen (onterecht) het gevoel dat het denkwerk voor hen gedaan wordt. Dit verleidt de gebruiker tot oppervlakkig cognitief gedrag dat het maken van gevolgtrekkingen en kennisvorming ontmoedigt. Gebruikers die informatie zelf internaliseren vertonen meer plangebaseerd gedrag, doen meer inspanning voor cognitieve processen, en zijn proactiever en bereid gevolgtrekkingen te maken. Dit zorgt voor meer focus, directere oplossingen, en betere opslag van de taakgerelateerde kennis. Deze kennis is gemakkelijker op te roepen in een toekomstige situatie. Ook is deze kennis beter overdraagbaar en toepasbaar in transfer situaties waar de taak, de interface, of beide anders zijn, en minder vatbaar voor een drastische taak onderbreking.

Betreffende de combinatie van theoretische kaders die we gebruiken concluderen we dat CLT waardevol is voor HCI, maar op zichzelf niet het hele scala aan cognitieve processen die plaatsvinden kan verklaren, net zoals Informatie Verwerkings Psychologie dat ook niet kan. CLT als theoretisch concept is waardevol om processen te beschrijven die plaatsvinden in de periferie van probleem oplossen in computer taken. De informationele processen zelf waar een probleem oplosser mee te maken krijgt, zouden echter zonder begrippen uit de IVP niet compleet beschreven kunnen worden. Begrippen uit CLT zoals Extraneous en Germane Cognitive Load kunnen gebruikt worden om te beschrijven hoe interface-gerelateerde gebeurtenissen ons gedrag kunnen beïnvloeden, in het bijzonder daar waar het motivatie en bereidheid tot inspanning betreft.

**Conclusies op het praktische niveau**

We concluderen dat waar het HCI richtlijnen betreft, er meer te overwegen valt dan men nu vaak doet, en men voordeel kan hebben van interfaces die ontworpen zijn volgens overwegingen die verder gaan dan “gewone gebruiksvriendelijkheid”. Onze experimenten betroffen een speciaal type van Externalisatie dat standaard is in hedendaagse interfaces: destination feedback. Het ontlasten van het werkgeheugen van een gebruiker en assisteren tijdens interactie door middel van destination feedback had negatieve consequenties. Het zorgde ervoor dat gebruikers rekenden op de interface, het gaf hen onterecht het gevoel dat het denkwerk gedeeltelijk voor hen werd gedaan, en verleidde hen tot oppervlakkiger cognitief gedrag. Bij het ontwerp van interfaces moeten we voorzichtig zijn met het aanbieden van hulp in de interface die teveel informatie weggeeft. We moeten zodanig ontwerpen dat de manier waarop gebruikers (zouden moeten) denken ondersteund wordt, en dat de software op deze manier zijn doel bereikt. We hebben een interactie-raamwerk voorgesteld dat beslissingen kan sturen bij interfaces die verder gaan dan taken waar alleen “gewone gebruiksvriendelijkheid” belangrijk is. Het kan richting geven bij interface ontwerp in serieuzere contexten, zoals onderwijssituaties waar leren zelf het doel is, of situaties waar veiligheid op het spel staat en continue proactiviteit van de gebruiker noodzakelijk is. Het kan ook belangrijk zijn wanneer het leunen op een interface ongewenst is, bijvoorbeeld bij risicovolle en complexe taken, of wanneer er de kans is dat een gebruiker plotseling met een nieuwe situatie geconfronteerd wordt waar bepaalde kennis toegepast moet worden. De algemene richtlijn “geef gebruikers niet de kans fouten te maken” moet natuurlijk niet genegeerd worden, maar tegelijkertijd zou de manier waarop de interactie plaatsvindt het leren van onderliggende regels moeten faciliteren. Hetzelfde geldt in situaties waar onderbrekingen normaal zijn, waar beheersing van het onderliggende gewenst is, of wanneer er kosten aan acties verbonden zijn en efficiënte oplossingen het doel zijn.

**Algemene conclusies**

Wanneer men de Externalisatie interface gebruikte, was taakprestatie op het moment zelf, en op een later tijdstip slechter dan wanneer men de Internalisatie interface gebruikte. Ook transfer of skill naar een andere taak en naar een andere interface was slechter met de Externalisatie interface. Deze laatste gebruikers vertoonden schermgebaseerd gedrag. Personen die de Internalisatie interface gebruikten, sloegen relevante taak- en regelkennis beter op in hun geheugen, en een drastische taakonderbreking had geen effect op hun prestatie (bij de Externalisatie personen wel). We maken hier uit op dat als men informatie zelf internaliseert, men meer plangebaseerd te werk gaat en meer bereid is zich in te spannen en gevolgtrekkingen te maken. Dit zorgt op zijn beurt voor een betere focus, directere en efficiëntere oplossingen, en betere opslag van taakkennis in het geheugen. Deze kennis is gemakkelijker terug te halen, op een toekomstig tijdstip, is minder gevoelig voor een drastische taakonderbreking, en is beter overdraagbaar naar en toepasbaar in transfer situaties. HCI ontwerpers kunnen hun voordeel doen met deze overwegingen die verder gaan dan “gewone gebruiksvriendelijkheid” richtlijnen, zelfs wanneer ze soms paradoxaal kunnen lijken.

## Curriculum Vitae

Christof van Nimwegen was born on August 9th 1971 in Nürnberg, Germany. He completed his primary and secondary education in Oosterhout and Breda, the Netherlands, after which he started studying Interaction Design at the Arts Academy in Utrecht, The Netherlands. After that, and a brief excursion to Artificial Intelligence, he studied Cognitive Psychology at Utrecht University, The Netherlands and received his master degree in Cognitive Ergonomics. Following his graduation, he worked several years as a usability engineer and interaction designer in Internet related businesses in the Netherlands and abroad. At the end of 2002, he became a junior teacher at Utrecht University, after which he enrolled in a Ph.D. project in 2003, concerning representations in interfaces, and Human Computer Interaction in general. Currently, Christof works as a senior researcher at the Center of Usability Research at the Catholic University Leuven (Belgium).

## List of publications

Although this thesis is not a bundle of published papers, several publications have appeared during this research project, and much of the findings reported on in the chapters can be found in them. Below is a list mentioning which chapters are based on, or (partly) inspired by which publications.

**Chapter 3**

Van Nimwegen, C., Van Oostendorp, H., & Tabachneck-Schijf, H. J. M. (2004a). Can more help be worse? The over-assisting interface. In Proceedings of SIGCHI.NL ’05: *Conference on Dutch directions in HCI*. New York, NY: ACM Press.

Van Nimwegen, C., Van Oostendorp, H., & Tabachneck-Schijf, H. J. M. (2004b). Does context sensitivity in the interface help? In proceedings of ECCE ’04: *The Twelfth European Conference on Cognitive Ergonomics*. France: European Association of Cognitive Ergonomics.

Van Nimwegen, C., Van Oostendorp, H., & Tabachneck-Schijf, H. J. M. (2004c). Externalization vs. Internalization: The Influence on Problem Solving Performance. In Proceedings of ICALT ’04: *The 4th IEEE Internationat. Conference on Advanced Learning Technologies*. Los Alamitos, CA: IEEE.

Van Nimwegen, C., Van Oostendorp, H., & Tabachneck-Schijf, H. J. M. (2004d). The over-assisting interface: The role of support in knowledge acquisition. In Proceedings of HCI2004: *Design for Life*. Bristol, UK: Research Press International (on behalf of the British HCI).

Van Nimwegen, C., Van Oostendorp, H., & Tabachneck-Schijf, H. J. M. (2005). The role of interface style in planning during problem solving. In Proceedings of CogSci ’05*: The 27th Annual Cognitive Science Conference*. Mahwah, NJ: Lawrence Erlbaum.

Van Nimwegen, C., Van Oostendorp, H., & Tabachneck-Schijf, H. J. M. (2006). Guidance in the interface: Effects of externalizing information during problem solving. In E. M. Alkhalifa (Ed.), *Cognitively Informed Systems: Utilizing Practical Approaches to Enrich Information Presentation and Transfer* (pp. 74-101). Hershey, PA: Idea Group Inc.

**Chapter 4**

Van Nimwegen, C., Van Oostendorp, H., Burgos, D., & Koper, R. (2006). Does an interface with less assistance provoke more thoughtful behavior? In Proceedings of ICLS ’06: *The 7th International Conference on Learning Sciences*. Bloomington, IN: International Society of the Learning Sciences.

Van Nimwegen, C., Burgos, D., Van Oostendorp, H., & Tabachneck-Schijf, H. J. M. (2006). The Paradox of the Assisted User: Guidance can be Counterproductive. In Proceedings of CHI’06: *The SIGCHI Conference on Human Factors in Computing Systems*. New York, NY: ACM.

Van Nimwegen, C., & Burgos, D. (2005). Performance in a planning task: The (ir)relevance of interface style and user's cognitive style. In Proceedings of SIGCHI.NL ‘05: *HCI Close 2U*. The Hague, The Netherlands.

Van Nimwegen, C., & Van Oostendorp, H. (2007a). Guidance in the interface and transfer of task performance. In proceedings of ECCE ’07: *The European Conference on Cognitive Ergonomics*. France: European Association of Cognitive Ergonomics.

Van Nimwegen, C., & Van Oostendorp, H. (2007b). Visual guidance and the impact of interruption during problem solving: Interface issues embedded in Cognitive Load Theory. In Proceedings of EuroCogSci ’07: *The Second European Cognitive Science Conference*. Mahwah, NJ: Lawrence Erlbaum Associates.

Van Nimwegen, C., & Van Oostendorp, H. (2008a). The questionable impact of an assisting interface on performance in transfer situations. *International Journal of Industrial Ergonomics* (under review).

Van Nimwegen, C., & Van Oostendorp, H. (2008b). Influencing Proactive User Behavior: Visual Assistance can be Counter-effective. *Interacting with Computers* (submitted).

**Chapter 5**

Van Nimwegen, C., & Van Oostendorp, H. (2006). Visual guidance in the interface and mouse movement behavior. Do we look where the relevant information is? In Proceedings of InScit ’07: *Current Research in Information Sciences and Technologies*. Merida, Spain: Open Institute of Knowledge.

**Chapter 6**

Burgos, D., Van Nimwegen, C., Van Oostendorp, H., & Koper, R. (2008). Game-based learning and the role of feedback. A case study of The Planning Educational Task. *International Journal of Advanced Technology for Learning (in press).*

## SIKS dissertation series

| | |
|---|---|
| 1998-1 | Johan van den Akker (CWI): DEGAS - An Active, Temporal Database of Autonomous Objects |
| 1998-2 | Floris Wiesman (UM): Information Retrieval by Graphically Browsing Meta-Information |
| 1998-3 | Ans Steuten (TUD): A Contribution to the Linguistic Analysis of Business Conversations within the Language/Action Perspective |
| 1998-4 | Dennis Breuker (UM): Memory versus Search in Games |
| 1998-5 | E.W.Oskamp (RUL): Computerondersteuning bij Straftoemeting |
| 1999-1 | Mark Sloof (VU): Physiology of Quality Change Modelling; Automated modelling of Quality Change of Agricultural Products |
| 1999-2 | Rob Potharst (EUR): Classification using decision trees and neural nets |
| 1999-3 | Don Beal (UM): The Nature of Minimax Search |
| 1999-4 | Jacques Penders (UM): The practical Art of Moving Physical Objects |
| 1999-5 | Aldo de Moor (KUB): Empowering Communities: A Method for the Legitimate User-Driven Specification of Network Information Systems |
| 1999-6 | Niek J.E. Wijngaards (VU): Re-design of compositional systems |
| 1999-7 | David Spelt (UT): Verification support for object database design |
| 1999-8 | Jacques H.J. Lenting (UM): Informed Gambling: Conception and Analysis of a Multi-Agent Mechanism for Discrete Reallocation. |
| 2000-1 | Frank Niessink (VU): Perspectives on Improving Software Maintenance |
| 2000-2 | Koen Holtman (TUE): Prototyping of CMS Storage Management |
| 2000-3 | Carolien M.T. Metselaar (UVA): Sociaal-organisatorische gevolgen van kennistechnologie; een procesbenadering en actorperspectief |
| 2000-4 | Geert de Haan (VU): ETAG, A Formal Model of Competence Knowledge for User Interface Design |
| 2000-5 | Ruud van der Pol (UM): Knowledge-based Query Formulation in Information Retrieval |
| 2000-6 | Rogier van Eijk (UU): Programming Languages for Agent Communication |
| 2000-7 | Niels Peek (UU): Decision-theoretic Planning of Clinical Patient Management |
| 2000-8 | Veerle Coupé (EUR): Sensitivity Analyis of Decision-Theoretic Networks |
| 2000-9 | Florian Waas (CWI): Principles of Probabilistic Query Optimization |
| 2000-10 | Niels Nes (CWI): Image Database Management System Design Considerations, Algorithms and Architecture |
| 2000-11 | Jonas Karlsson (CWI): Scalable Distributed Data Structures for Database Management |
| 2001-1 | Silja Renooij (UU): Qualitative Approaches to Quantifying Probabilistic Networks |
| 2001-2 | Koen Hindriks (UU): Agent Programming Languages: Programming with Mental Models |
| 2001-3 | Maarten van Someren (UvA): Learning as problem solving |
| 2001-4 | Evgueni Smirnov (UM): Conjunctive and Disjunctive Version Spaces with Instance-Based Boundary Sets |

| | |
|---|---|
| 2001-5 | Jacco van Ossenbruggen (VU): Processing Structured Hypermedia: A Matter of Style |
| 2001-6 | Martijn van Welie (VU): Task-based User Interface Design |
| 2001-7 | Bastiaan Schonhage (VU): Diva: Architectural Perspectives on Information Visualization |
| 2001-8 | Pascal van Eck (VU): A Compositional Semantic Structure for Multi-Agent Systems Dynamics. |
| 2001-9 | Pieter Jan 't Hoen (RUL): Towards Distributed Development of Large Object-Oriented Models, Views of Packages as Classes |
| 2001-10 | Maarten Sierhuis (UvA): Modeling and Simulating Work Practice BRAHMS: a multiagent modeling and simulation language for work practice analysis and design |
| 2001-11 | Tom M. van Engers (VUA): Knowledge Management: The Role of Mental Models in Business Systems Design |
| 2002-01 | Nico Lassing (VU): Architecture-Level Modifiability Analysis |
| 2002-02 | Roelof van Zwol (UT): Modelling and searching web-based document Collections |
| 2002-03 | Henk Ernst Blok (UT): Database Optimization Aspects for Information Retrieval |
| 2002-04 | Juan Roberto Castelo Valdueza (UU): The Discrete Acyclic Digraph Markov Model in Data Mining |
| 2002-05 | Radu Serban (VU): The Private Cyberspace Modeling Electronic Environments inhabited by Privacy-concerned Agents |
| 2002-06 | Laurens Mommers (UL): Applied legal epistemology; Building a knowledge-based ontology of the legal domain |
| 2002-07 | Peter Boncz (CWI): Monet: A Next-Generation DBMS Kernel For Query-Intensive Applications |
| 2002-08 | Jaap Gordijn (VU): Value Based Requirements Engineering: Exploring Innovative E-Commerce Ideas |
| 2002-09 | Willem-Jan van den Heuvel( KUB): Integrating Modern Business Applications with Objectified Legacy Systems |
| 2002-10 | Brian Sheppard (UM): Towards Perfect Play of Scrabble |
| 2002-11 | Wouter C.A. Wijngaards (VU): Agent Based Modelling of Dynamics: Biological and Organisational Applications |
| 2002-12 | Albrecht Schmidt (Uva): Processing XML in Database Systems |
| 2002-13 | Hongjing Wu (TUE): A Reference Architecture for Adaptive Hypermedia Applications |
| 2002-14 | Wieke de Vries (UU): Agent Interaction: Abstract Approaches to Modelling, Programming and Verifying Multi-Agent Systems |
| 2002-15 | Rik Eshuis (UT): Semantics and Verification of UML Activity Diagrams for Workflow Modelling |
| 2002-16 | Pieter van Langen (VU): The Anatomy of Design: Foundations, Models and Applications |
| 2002-17 | Stefan Manegold (UVA): Understanding, Modeling, and Improving Main-Memory Database Performance |

2003-01 Heiner Stuckenschmidt (VU): Ontology-Based Information Sharing in Weakly Structured Environments
2003-02 Jan Broersen (VU): Modal Action Logics for Reasoning About Reactive Systems
2003-03 Martijn Schuemie (TUD): Human-Computer Interaction and Presence in Virtual Reality Exposure Therapy
2003-04 Milan Petkovic (UT): Content-Based Video Retrieval Supported by Database Technology
2003-05 Jos Lehmann (UVA): Causation in Artificial Intelligence and Law - A modelling approach
2003-06 Boris van Schooten (UT): Development and specification of virtual environments
2003-07 Machiel Jansen (UvA): Formal Explorations of Knowledge Intensive Tasks
2003-08 Yongping Ran (UM): Repair Based Scheduling
2003-09 Rens Kortmann (UM): The resolution of visually guided behaviour
2003-10 Andreas Lincke (UvT): Electronic Business Negotiation: Some experimental studies on the interaction between medium, innovation context and culture
2003-11 Simon Keizer (UT): Reasoning under Uncertainty in Natural Language Dialogue using Bayesian Networks
2003-12 Roeland Ordelman (UT): Dutch speech recognition in multimedia information retrieval
2003-13 Jeroen Donkers (UM): Nosce Hostem - Searching with Opponent Models
2003-14 Stijn Hoppenbrouwers (KUN): Freezing Language: Conceptualisation Processes across ICT-Supported Organisations
2003-15 Mathijs de Weerdt (TUD): Plan Merging in Multi-Agent Systems
2003-16 Menzo Windhouwer (CWI): Feature Grammar Systems - Incremental Maintenance of Indexes to Digital Media Warehouses
2003-17 David Jansen (UT): Extensions of Statecharts with Probability, Time, and Stochastic Timing
2003-18 Levente Kocsis (UM): Learning Search Decisions
2004-01 Virginia Dignum (UU): A Model for Organizational Interaction: Based on Agents, Founded in Logic
2004-02 Lai Xu (UvT): Monitoring Multi-party Contracts for E-business
2004-03 Perry Groot (VU): A Theoretical and Empirical Analysis of Approximation in Symbolic Problem Solving
2004-04 Chris van Aart (UVA): Organizational Principles for Multi-Agent Architectures
2004-05 Viara Popova (EUR): Knowledge discovery and monotonicity
2004-06 Bart-Jan Hommes (TUD): The Evaluation of Business Process Modeling Techniques
2004-07 Elise Boltjes (UM): Voorbeeldig onderwijs; voorbeeldgestuurd onderwijs, een opstap naar abstract denken, vooral voor meisjes
2004-08 Joop Verbeek (UM): Politie en de Nieuwe Internationale Informatiemarkt, Grensregionale politiële gegevensuitwisseling en digitale expertise
2004-09 Martin Caminada (VU): For the Sake of the Argument; explorations into argument-based reasoning

| | |
|---|---|
| 2004-10 | Suzanne Kabel (UVA): Knowledge-rich indexing of learning-objects |
| 2004-11 | Michel Klein (VU): Change Management for Distributed Ontologies |
| 2004-12 | The Duy Bui (UT): Creating emotions and facial expressions for embodied agents |
| 2004-13 | Wojciech Jamroga (UT): Using Multiple Models of Reality: On Agents who Know how to Play |
| 2004-14 | Paul Harrenstein (UU): Logic in Conflict. Logical Explorations in Strategic Equilibrium |
| 2004-15 | Arno Knobbe (UU): Multi-Relational Data Mining |
| 2004-16 | Federico Divina (VU): Hybrid Genetic Relational Search for Inductive Learning |
| 2004-17 | Mark Winands (UM): Informed Search in Complex Games |
| 2004-18 | Vania Bessa Machado (UvA): Supporting the Construction of Qualitative Knowledge Models |
| 2004-19 | Thijs Westerveld (UT): Using generative probabilistic models for multimedia retrieval |
| 2004-20 | Madelon Evers (Nyenrode): Learning from Design: facilitating multidisciplinary design teams |
| 2005-01 | Floor Verdenius (UVA): Methodological Aspects of Designing Induction-Based Applications |
| 2005-02 | Erik van der Werf (UM): AI techniques for the game of Go |
| 2005-03 | Franc Grootjen (RUN): A Pragmatic Approach to the Conceptualisation of Language |
| 2005-04 | Nirvana Meratnia (UT): Towards Database Support for Moving Object data |
| 2005-05 | Gabriel Infante-Lopez (UVA): Two-Level Probabilistic Grammars for Natural Language Parsing |
| 2005-06 | Pieter Spronck (UM): Adaptive Game AI |
| 2005-07 | Flavius Frasincar (TUE): Hypermedia Presentation Generation for Semantic Web Information Systems |
| 2005-08 | Richard Vdovjak (TUE): A Model-driven Approach for Building Distributed Ontology-based Web Applications |
| 2005-09 | Jeen Broekstra (VU): Storage, Querying and Inferencing for Semantic Web Languages |
| 2005-10 | Anders Bouwer (UVA): Explaining Behaviour: Using Qualitative Simulation in Interactive Learning Environments |
| 2005-11 | Elth Ogston (VU): Agent Based Matchmaking and Clustering - A Decentralized Approach to Search |
| 2005-12 | Csaba Boer (EUR): Distributed Simulation in Industry |
| 2005-13 | Fred Hamburg (UL): Een Computermodel voor het Ondersteunen van Euthanasiebeslissingen |
| 2005-14 | Borys Omelayenko (VU): Web-Service configuration on the Semantic Web; Exploring how semantics meets pragmatics |
| 2005-15 | Tibor Bosse (VU): Analysis of the Dynamics of Cognitive Processes |
| 2005-16 | Joris Graaumans (UU): Usability of XML Query Languages |
| 2005-17 | Boris Shishkov (TUD): Software Specification Based on Re-usable Business Component |

2005-18 Danielle Sent (UU): Test-selection strategies for probabilistic networks
2005-19 Michel van Dartel (UM): Situated Representation
2005-20 Cristina Coteanu (UL): Cyber Consumer Law, State of the Art and Perspectives
2005-21 Wijnand Derks (UT): Improving Concurrency and Recovery in Database Systems by Exploiting Application Semantics
2006-01 Samuil Angelov (TUE): Foundations of B2B Electronic Contracting
2006-02 Cristina Chisalita (VU): Contextual issues in the design and use of information technology in organizations
2006-03 Noor Christoph (UVA): The role of metacognitive skills in learning to solve problems
2006-04 Marta Sabou (VU): Building Web Service Ontologies
2006-05 Cees Pierik (UU): Validation Techniques for Object-Oriented Proof Outlines
2006-06 Ziv Baida (VU): Software-aided Service Bundling - Intelligent Methods & Tools for Graphical Service Modeling
2006-07 Marko Smiljanic (UT): XML schema matching -- balancing efficiency and effectiveness by means of clustering
2006-08 Eelco Herder (UT): Forward, Back and Home Again - Analyzing User Behavior on the Web
2006-09 Mohamed Wahdan (UM): Automatic Formulation of the Auditor's Opinion
2006-10 Ronny Siebes (VU): Semantic Routing in Peer-to-Peer Systems
2006-11 Joeri van Ruth (UT): Flattening Queries over Nested Data Types
2006-12 Bert Bongers (VU): Interactivation - Towards an e-cology of people, our technological environment, and the arts
2006-13 Henk-Jan Lebbink (UU): Dialogue and Decision Games for Information Exchanging Agents
2006-14 Johan Hoorn (VU): Software Requirements: Update, Upgrade, Redesign – towards a Theory of Requirements Change
2006-15 Rainer Malik (UU): CONAN: Text Mining in the Biomedical Domain
2006-16 Carsten Riggelsen (UU): Approximation Methods for Efficient Learning of Bayesian Networks
2006-17 Stacey Nagata (UU): User Assistance for Multitasking with Interruptions on a Mobile Device
2006-18 Valentin Zhizhkun (UVA): Graph transformation for Natural Language Processing
2006-19 Birna van Riemsdijk (UU): Cognitive Agent Programming: A Semantic Approach
2006-20 Marina Velikova (UvT): Monotone models for prediction in data mining
2006-21 Bas van Gils (RUN): Aptness on the Web
2006-22 Paul de Vrieze (RUN): Fundaments of Adaptive Personalisation
2006-23 Ion Juvina (UU): Development of Cognitive Model for Navigating on the Web
2006-24 Laura Hollink (VU): Semantic Annotation for Retrieval of Visual Resources
2006-25 Madalina Drugan (UU): Conditional log-likelihood MDL and Evolutionary MCMC
2006-26 Vojkan Mihajlovic (UT): Score Region Algebra: A Flexible Framework for

Structured Information Retrieval
2006-27 Stefano Bocconi (CWI): Vox Populi: generating video documentaries from semantically annotated media repositories
2006-28 Borkur Sigurbjornsson (UVA): Focused Information Access using XML Element Retrieval
2007-01 Kees Leune (UvT): Access Control and Service-Oriented Architectures
2007-02 Wouter Teepe (RUG): Reconciling Information Exchange and Confidentiality: A Formal Approach
2007-03 Peter Mika (VU): Social Networks and the Semantic Web
2007-04 Jurriaan van Diggelen (UU): Achieving Semantic Interoperability in Multi-agent Systems: a dialogue-based approach
2007-05 Bart Schermer (UL): Software Agents, Surveillance, and the Right to Privacy: a Legislative Framework for Agent-enabled Surveillance
2007-06 Gilad Mishne (UVA): Applied Text Analytics for Blogs
2007-07 Natasa Jovanovic' (UT): To Whom It May Concern - Addressee Identification in Face-to-Face Meetings
2007-08 Mark Hoogendoorn (VU): Modeling of Change in Multi-Agent Organizations
2007-09 David Mobach (VU): Agent-Based Mediated Service Negotiation
2007-10 Huib Aldewereld (UU): Autonomy vs. Conformity: an Institutional Perspective on Norms and Protocols
2007-11 Natalia Stash (TUE): Incorporating Cognitive/Learning Styles in a General-Purpose Adaptive Hypermedia System
2007-12 Marcel van Gerven (RUN): Bayesian Networks for Clinical Decision Support: A Rational Approach to Dynamic Decision-Making under Uncertainty
2007-13 Rutger Rienks (UT): Meetings in Smart Environments; Implications of Progressing Technology
2007-14 Niek Bergboer (UM): Context-Based Image Analysis
2007-15 Joyca Lacroix (UM): NIM: a Situated Computational Memory Model
2007-16 Davide Grossi (UU): Designing Invisible Handcuffs. Formal investigations in Institutions and Organizations for Multi-agent Systems
2007-17 Theodore Charitos (UU): Reasoning with Dynamic Networks in Practice
2007-18 Bart Orriens (UvT): On the development an management of adaptive business collaborations
2007-19 David Levy (UM): Intimate relationships with artificial partners
2007-20 Slinger Jansen (UU): Customer Configuration Updating in a Software Supply Network
2007-21 Karianne Vermaas (UU): Fast diffusion and broadening use: A research on residential adoption and usage of broadband internet in the Netherlands between 2001 and 2005
2007-22 Zlatko Zlatev (UT): Goal-oriented design of value and process models from patterns
2007-23 Peter Barna (TUE): Specification of Application Logic in Web Information Systems

| | |
|---|---|
| 2007-24 | Georgina Ramírez Camps (CWI): Structural Features in XML Retrieval |
| 2007-25 | Joost Schalken (VU): Empirical Investigations in Software Process Improvement |
| 2008-01 | Katalin Boer-Sorbán (EUR): Agent-Based Simulation of Financial Markets: A modular, continuous-time approach |
| 2008-02 | Alexei Sharpanskykh (VU): On Computer-Aided Methods for Modeling and Analysis of Organizations |
| 2008-03 | Vera Hollink (UVA): Optimizing hierarchical menus: a usage-based approach |
| 2008-04 | Ander de Keijzer (UT): Management of Uncertain Data - towards unattended integration |
| 2008-05 | Bela Mutschler (UT): Modeling and simulating causal dependencies on process-aware information systems from a cost perspective |
| 2008-06 | Arjen Hommersom (RUN): On the Application of Formal Methods to Clinical Guidelines, an Artificial Intelligence Perspective |
| 2008-07 | Peter van Rosmalen (OU): Supporting the tutor in the design and support of adaptive e-learning |
| 2008-08 | Janneke Bolt (UU): Bayesian Networks: Aspects of Approximate Inference |
| 2008-09 | Christof van Nimwegen (UU): The paradox of the guided user: assistance can be counter-effective |