Can more help be worse? The over-assisting interface

Christof van Nimwegen Herre van Oostendorp Hermina (Tabachneck-) Schijf

Center for Content and Knowledge Engineering, Utrecht University

Padualaan 14, 3584CH Utrecht, The Netherlands

{christof, herre, h.schijf}@cs.uu.nl

ABSTRACT

The premise of this research is that assistance can make interactions easier, but does not necessarily aid understanding. A way to assist users is displaying relevant information in interfaces. Recall of such knowledge is unnecessary and working memory is relieved. Examples are visual aids such as the “greying-out” of items, which can become unclickable indicating that an action is not possible. If this is not done this information has to be inferred by the user himself. An experiment was conducted in which subjects had to solve a series of puzzles. We hypothesized that providing greyed-out items (externalization) yields better performance during initial learning. An interface without greying-out (internalization) is expected to yield better performance in later phases, and better knowledge of the task. Subjects solved an isomorph of “missionaries and cannibals” in two conditions: with greyed-out items and without. It showed that externalization had little influence on performance. All subjects learned quite well how to solve the puzzle. On a knowledge test however, it turned out to be different. The procedural knowledge tested afterwards was equal, but declarative knowledge, concerning the rule central to what this problem was about, was worse for persons who had greyed-out items. Also, months later the same internalization-subjects had faster problem recognition of the task, and better performance on a similar task.

Keywords

Interface, screen representation, problem solving, planning, declarative knowledge, visualization

INTRODUCTION

The notion that learning is more effective when people experiment themselves is far from new, and exploratory learning as a whole has been a subject of research in many domains. Carroll [1] for example, already more than a decade ago propagated minimalism in design and instructions. Looking at software today, one cannot fail to see that it has seen a tremendous development during the last decades. Being more advanced, the amount, diversity and turnover demand fast learning from users.

LEAVE BLANK THE LAST 2.5 cm (1”) OF THE LEFT COLUMN ON THE FIRST PAGE FOR THE COPYRIGHT NOTICE.

Considerable accent has been put on usability (effectiveness, efficiency & satisfaction) that users experience [2]; interfaces of today are by no means comparable with the command-line interfaces it all began with. Interfaces often are very complex, and applications have hundreds if not thousands of functions. A notion that came with usability is the importance of “minimizing user memory load“. A recommendation that often was heard is that users should be able to interact on basis of recognition rather than recall [2]. In practice this means that objects, actions, and options should be visible, and the user should not have to remember too much information. In our daily interactions with software, we are nowadays quite used to see quite a portion of the interface controls greyed out (fig.1). This often is meant to “take the user by the hand” by limiting choices and providing feedback [5]. There are various reasons for greying-out buttons. In some “wizards” where a certain sequence of actions has to be performed, having a button greyed-out often indicates that one has forgotten to fill out a mandatory field. In other applications, greyed-out menu-items indicate that using the item is out of context.

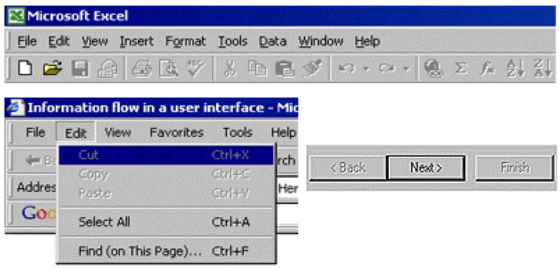

Figure 1 – Greyed out interface items

The interface is thus context-sensitive, and offers only possible actions. The “greying-out” that is applied is supposed to help the user, telling him/her that in the current situation applying a specific function is not allowed, and remembering this is not necessary. However, the user is left in the dark, as to the reason for the greying out. Some examples are easy to grasp. If nothing is copied, you cannot paste it, so “paste” is greyed-out, but reasons for greying-out are not always so self-evident. This research specifically focuses on externalizing vs. internalizing information. By externalizing certain information, working memory (WM) is relieved [6]. Several other studies showed that externalizing information can be useful for cognitive tasks; the more is externalized, the easier it is to solve the problem. This was generally accepted, but as Zhang and Wang [7] pointed out, there hadn't been research on how exactly external representation influences WM. They distinguished between IWM (internal representations & memory processes) and EWM (external representations & perceptual processes). It showed that having *all* information in EWM always aided problem solving, but with information distributed across IWM and EWM it could go both ways, enhancing or hindering performance. Of influence were retrieval strategy and the way information was encoded. The results are not clear-cut, and difficult to generalize and apply to interfaces and problem solving tasks. In computer programs, it is not always possible to have all functions externalized. To contribute to a more complete theory, research is needed on effects of distributing information. One could argue that with most information externalized (and little internalized) users are not triggered to look for underlying rules. In contrast, having most knowledge internalized is perhaps important when the task is interrupted, when transfer is needed, or when speeding up of tasks is important; users can fall back on internalized knowledge. O'Hara and Payne [3] and Trudel and Payne [4] provide support for this point of view, stating that too strong a reliance on external information leads to negative effects regarding planning and transfer of skills. They distinguished between plan-based and display-based problem solving. In plan-based problem solving one uses detailed problem strategies from long-term memory. Display-based makes little use of learned knowledge but relies on interface information. Plan-based activity leads to shorter solution routes (less unnecessary steps), because steps are planned, while display-based strategy involves more steps because of more searching. This research is part of a broader research program aiming to contribute to a theory that explains and predicts which type of screen representation leads to better task performance in terms of learning, task memory after delay, and transfer. Later we will test our hypotheses on realistic situated tasks.

HYPOTHESES

We tried to investigate our questions using the following hypotheses:

1. **Externalization will initially yield better task performance than internalization.** When internal knowledge-elements are not yet acquired, externalization will be helpful in the beginning
2. **Internalization yields better task performance in later phases.** In later phases internalization allows reliance on better-internalized knowledge, leading to better performance.
3. **Internalization yields better knowledge.** Not relying on externalized information provokes planning steps oneself, consequently knowledge of the task will be better

METHOD

Material

The problem-solving task we used is called "Balls & Boxes" (B&B). It is an informationally equivalent isomorph of "Missionaries and Cannibals" (M&C). 5 missionaries and 5 cannibals stand on a riverbank, and all have to reach the other bank by boat. The boat only fits 3 people, the minimum to sail is 1, and cannibals can never outnumber missionaries at any place, or the latter will be eaten (fig.4, but with 5 instead of 3 M&C). The B&B problem (fig.2) uses the same problem space, but is more abstract (in M&C rules are quite intuitive; cannibals eat people, boats cross rivers). Using boxes and blue and yellow balls and a dish instead, we avoided too easily learned rules. Rules are as follows:

1. Balls should be transported using the dish
2. You can transport at most 3 balls at a time
3. To move, the dish must contain at least 1 ball
4. The dish has to reach the boxes in turn
5. No more blue than yellow balls in the dish
6. No more blue than yellow balls left in the boxes

This is the kind of puzzle with a "trick" to it, an aha-moment from which point on the solution is at hand. The controls were simple: to get balls into the dish, blue or yellow up-arrows had to be clicked and to move the dish horizontally, one had to click a pink arrow (left or right). At all times, rules could be consulted as in the list below, by clicking a tab. There were 2 puzzle versions: With greying-out (Externalization): Arrows are only coloured (clickable) when an action is legal, and greyed out (unclickable) if illegal. Moving the dish empty in fig.2 is illegal because it violates rule 3. In this situation, it is externalized by greying out the pink arrows. Without greying-out (Internalization): All arrows are always coloured (clickable), providing no information about the legality of moves. One can push all buttons at all times. If one wants to move the dish empty (illegal) and clicks a pink arrow, the dish will move to the right, but an error notification pops up saying "this is not possible". Subjects had to click "ok" to undo the move.

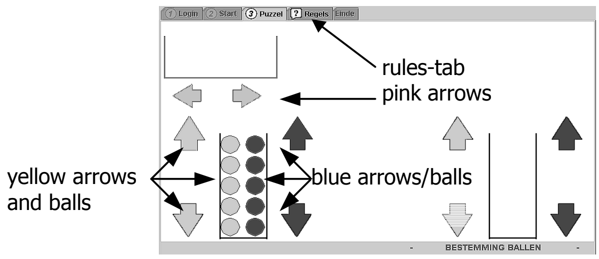

Figure 2 - The balls & boxes puzzle

Knowledge test

To estimate the knowledge of subjects, we developed a knowledge test of 8 questions. The puzzle has more than 30 legal states. There were 7 procedural knowledge questions about 7 of those states (open and multiple choice) in which subjects were presented with puzzle situations visually (fig.3). They had to judge whether or not certain moves were leading to the solution, and why.

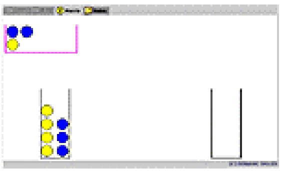

Figure 3 – Puzzle situation from knowledge test

To test declarative knowledge, there are limited possibilities. The puzzle had only 6 rules. Rules 1 to 4 were easy to grasp and remember. Rules 5 and 6 are more crucial, the ones that define the difficulty of this puzzle, and can actually be merged to just 1 rule: “blue balls can never be in the majority at any place except when there are only blue balls”. Subjects were asked about this rule with a multiple-choice question.

Material for delayed re-testing

To see what happens over time we re-invited the same subjects after 8 months. Again they received the B&B puzzle, but no rules were consultable, we purely wanted to check performance after a long delay. In addition, subjects solved a “real” Missionaries & Cannibals puzzle that was semantically richer (fig. 4). The solution (11 steps) is similar as in B&B. However, the fact that the playing direction was reversed and that there were 3 instead of 5 of each objects will force subjects to stop, think and apply (transfer) learned knowledge to this similar problem.

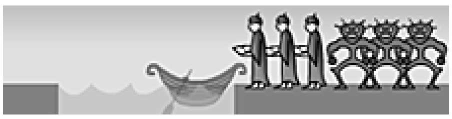

Figure 4 - The missionaries & cannibals puzzle

Procedure

Thirty subjects (15 per version) aged 19-28, experienced with PC’s had to solve the B&B puzzle 9 times. Always the starting situation differed, to avoid subjects relying too much on “having learned the trick” and repeat actions. For a delayed re-test 8 months later, we again invited 14 of the same subjects, 7 from each version. They had to solve the same B&B puzzle again (5 times) without further instructions. After this they had to solve the realistic M&C puzzle during 8 minutes.

RESULTS

Among the measures we analysed were the number of solved trials, as well as time and steps needed to solve the puzzle. On most measures the average score of internalization-subjects was higher, but not significantly so. In both groups, already after 4 trials on average, a ceiling was reached, and all the subjects regardless of the version they worked with, were perfectly able to perform fast and efficiently (the minimum of 11 steps). What we did find were nearly significant differences in terms of performance-behaviour. Internalization-subjects performed better, reached less dead-end states, they made fewer un-smart moves. As said, after 4 trials, few mistakes were made; “the trick” was mastered. We confirmed that there was no difference in rule-checking behaviour; in both versions they were consulted equally often, and as expected especially in the beginning.

The scores on the procedural knowledge questions for both externalization and internalization were quit high, 5.8 and 6.2 respectively (7 was the maximum). This small difference was not significant; both groups basically answered these questions equally well. For declarative knowledge however (fig.5), we found that subjects that had no greyed-out items (internalization) scored far better. All the internalization-subjects correctly answered this question, whereas of the externalization-subjects only 60% answered it correctly. This relationship was significant (Phi =-.50, $p < 0.01$).

Figure 5 – Proportion correct declarative knowledge per version

In the delayed-retest encouraging results were found. In the same B&B puzzle internalization-subjects needed significantly less time to solve it the first time (mean **432** sec, Sd 314 and mean **778** sec, Sd 397 resp; T (df=12)=-1.81, $p < .05$). After the first success, all subjects solved the puzzle equally well. Just as 8 months ago, the ceiling was again reached.

Figure 6 – Time (sec) until success B&B per version

Figure 7 – Number of M&C solved per version

In the second task (transfer) internalization-subjects managed to solve the M&C (fig.7) puzzle significantly more often than externalization-subjects (mean **4.14**, Sd 2.54 and mean **1.43**, Sd 1.81; T (df=12)= 2.30, $p < .05$).

DISCUSSION

Our first hypothesis stating that initially having greyed-out items (externalization) yields better performance was not supported (unlike in research of others). There were minor differences in time and steps needed in the predicted direction, but not significantly so.

The second hypotheses stating that not having greyed-out items (internalization) would yield better performance in later stages was partly supported. On the measure "reaching dead-end states" showing how subjects behaved, in terms of the insight they had some confirming results were found. The assumption that internalization-subjects do some smarter, more elaborated planning while externalization-subjects solve more on base of interface cues indeed seemed to be true. This difference was more visible in later phases. Internalization-subjects performed better, reaching those dead-end problem-states less. This confirms our expectation that internalization causes subjects to work on base of a more plan-based strategy. More interesting differences emerged after quite a longer delay, see further in the discussion.

What we want to focus on first is an interesting observation concerning the third hypothesis. We expected subjects that had no greyed-out items (internalization) to have better knowledge after the experiment, because they had to build a stronger, more elaborated plan that relied less on interface information. For procedural knowledge, we were surprised to find no differences. Being confronted with screenshots of puzzle states and asked to judge them, both groups had equally good knowledge. Was it because the questions were asked with a visual example and this caused recognition? Were these visual cues sufficient for subjects to trigger the needed procedural knowledge, or was the puzzle so easy? There seems to be more to it. Looking at declarative knowledge it came as a surprise that even though the mentioned performance and procedural knowledge were equally high in both groups, externalization-subjects were much worse at correctly recalling the actual rule the puzzle was about (60% vs. 100%). Just over half the subjects that used greyed-out items knew it correctly, opposed to 100% in the internalization version. They "did it right", but not based on correct declarative knowledge. Perhaps (procedural) knowledge they possessed here was comparable with the way people sometimes remember safe combinations or phone numbers; just when they do it. We interpret these findings as indicators of better knowledge representation instigated by internalization. More confirmation came after the delayed re-test. In context of hypothesis 2, after 8 months, the same subjects solved the same M&C again. After such a long time, of course all subjects had to think and remember, but the first success took subjects that had worked with greyed-out items (externalization) twice as long as the internalization subjects. This indicates that after this time-lapse there was a difference in procedural knowledge as well. After this first success, performance in both groups was again at top-level very soon, so we assume that procedural knowledge at that moment was also equal again. Right after this moment, we confronted the subjects with another more realistic M&C puzzle that was actually the same problem in another appearance.

The subjects that had no greyed-out items 8 months before still had an advantage. This M&C puzzle was solved almost 3 times more often in the same time by internalization-subjects. We interpret this in a transfer-context; facing this "new" situation makes internalization-subjects understand better that this was actually the same problem.

Some remarks remain. Perhaps internalization-subjects might have done better in the initial B&B-puzzle if the application hadn't forced unintentional delays on them. They had to click away messages, and the program undid "wrong" moves. Maybe they lost time in recovering, described by O'Hara and Payne [3] as the effect of "lockout time and error recovery cost". It can also be that this particular puzzle was easy, and/or too prone to trial and error solving.

These experiments are a pilot study with rather abstract material meant to figure out if there is something to our notion that externalization might aid interaction, but that it does not necessarily aid understanding. The results were encouraging, and to let outcomes of future research contribute to GUI design guidelines, we will continue the research with more difficult problems and more realistic planning-related tasks, e.g. spreadsheet applications or drawing applications with less repetitive and more complex tasks.

REFERENCES

1. Carrol, J.M. The Nurnberg Funnel: Designing Minimalist Instruction for Practical Computer Skill. MIT Press, Cambridge, Mass, 1990.

2. Nielsen, J. Usability Engineering. Morgan Kaufmann, San Francisco, 1990.

3. O'Hara, K. and Payne, S.J. Planning and the user interface: the effects of lockout time and error recovery cost. *International Journal of Human-Computer Studies 50*, 41-49, 1999.

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

5. Van Oostendorp, H. and De Mul, S. Learning by exploration: Thinking aloud while exploring an information system. *Instructional Science 27*, 269-284, 1999.

6. Zhang, J. The nature of external representations in problem solving". *Cognitive Science 21,* 2, 179-217, 1997.

7. Zhang, J. and Wang, H. An Exploration of the Relations Between External Representations and Working Memory. (in review), 1998.