Does context sensitivity in the interface help?

Christof van Nimwegen Herre van Oostendorp Hermina J.M.(Tabachneck-) Schijf
Institute of Information and Computing Sciences
Center for Content and Knowledge Engineering, Utrecht University
Padualaan 14, 3584 Utrecht, The Netherlands
{christof, herre, h.schijf}@cs.uu.nl

ABSTRACT
A common phenomenon in graphical interfaces is to have some degree of context sensitivity. Providing context sensitive feedback, thus making recall of certain knowledge unnecessary and relieving working memory, is a means of externalizing information. Examples are visual feedback aids such as “greying out” (menu)-items. With internalization, certain task-information is less directly available, and needs to be internalized and stored in memory. We conducted an experiment investigating the effects of externalization vs. internalization on performance and knowledge acquisition. In 2 conditions subjects solved an isomorph of “missionaries and cannibals” (internalized or externalized). After this, knowledge of the problem’s rules was tested. Internalization resulted in better knowledge and reaching fewer dead-end problem-states. Months later, internalization-subjects had faster problem recognition of the task, and better performance on a similar task (transfer). This research contributes to HCI design-guidelines for problem-solving tasks.

Keywords
Interface, visualization, screen representation, problem solving, planning, externalization of information

INTRODUCTION
During the last decades, software has seen a tremendous development. Becoming more advanced, the amount, diversity and turnover demand fast learning from users. Accents have been put on usability (effectiveness, efficiency & satisfaction) that users experience (Nielsen, 1994); interfaces of today are by no means comparable with command-line interfaces it began with. Nowadays, a tendency is to “take the user by the hand”, by limiting choices, and providing feedback (Van Oostendorp & De Mul, 1999). Examples are wizards, help-options and greying-out menu-items that don’t permit using them, thus offering a context-sensitive interface with only “possible” actions (fig.1).

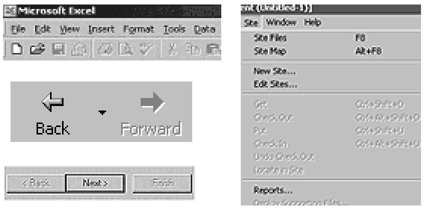

Figure 1 – Interface examples of greying out

For example in Word, you cannot select “paste” from the “edit”-tab in the menu, when nothing is copied first. “Paste” shown in grey colour indicates its presence, but using it is impossible. “Greying out” is an example of **externalizing** information, indicating that in that situation applying a specific function is not allowed. By making information available on the interface, remembering certain information is not necessary. By externalizing certain information, working memory (WM) is relieved (Zhang, 1997). 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 (1998) pointed out, there hadn’t been research on how exactly external representation influences WM. They designed a framework of Distributed WM used during cognitive tasks based on the WM Framework (Baddeley & Hitch, 1974). It has two components: IWM (internal representations & memory processes) and EWM (external representations & perceptual processes). In their experiments the issue was recalling series of digits. They showed that having *all* information in EWM always aided problem solving, but when information was 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, so also here we need a way to distribute information over IWM and EWM effectively. To contribute to a more complete theory, research is needed on effects of distributing information in different ways on performance. One could argue that most information externalized (and little internalized) seduces users not to reason – they are not triggered to look for underlying rules, whilst those might be necessary in order to build up stable knowledge that also can be applied in new situations. Perhaps having most knowledge internalized is also important when the task is interrupted, when dependence on a particular interface is not desired, when transfer is needed, or when speeding up of tasks is important. O’Hara and Payne (1999) and Trudel and Payne (1996) 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 drew a distinction 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, because steps are planned, while a display-based strategy involves more steps because of more searching. Perhaps plan-based behaviour results in still other favourable outcomes. A measure to look at could be the avoidance of "dead-end states", far from the solution, from where the only thing to do is go back. The research reported on here is part of a broader research programme entailing the testing of the hypotheses mentioned. We aim to contribute to a theory that explains and predicts which type of screen representation leads to better task performance in terms of learning, performance after learning, memory for the task after delay, and transfer. In later phases we will test our hypotheses on more realistic situated tasks.

HYPOTHESES

We tried to investigate our questions using the following hypotheses (fig.2):

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 later after a distraction task.** After interruption, internalization allows relying on better internalized knowledge, leading to better performance.
3. **Internalization yields better knowledge.** Not relying on externalized information provokes planning steps oneself. Knowledge of rules will be better here; this can also help with transfer.

We conducted an experiment in two sessions. In session 1, subjects solved a problem 9 times on a PC in 2 conditions: internalization and externalization. We expected the following outcomes:

Phase 1: 3 trials. Subjects don't have needed knowledge available yet, externalization offers more interface cues, and allows better performances.

Phase 2: 3 trials. Performance in the 2 conditions will become more or less equal, since internalization-subjects by then will have acquired internal knowledge.

Interruption: A distraction task erasing WM

Phase 3: 3 trials. Internalization-subjects perform better because of better-internalized knowledge and more elaborated planning.

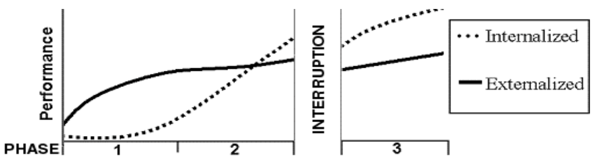

Figure 2 – Hypotheses

Because we expected internalization to result in better knowledge and stronger memory traces, we conducted a second session 8 months later. We hypothesized that internalization-subjects would still remember better how to solve the problem, and also perform better on a similar task when transfer occurs.

METHOD

Material Session 1

The problem-solving task we used in session 1 is called "Balls & Boxes" (B&B), an isomorphic version 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. Constraints are that the boat only fits 3, the minimum to sail is 1, and cannibals cannot outnumber missionaries at any place, or the latter will be eaten (fig.5, with 5 instead of 3 M&C). The B&B problem (fig. 3) uses the same problem space, but is more abstract, because in M&C rules are easy to learn (cannibals eat people, boats cross rivers etc.). Using boxes and blue and yellow balls and a dish instead, we avoided too easily learned rules. The rules were:

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 boxes

The puzzle worked quite 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). Consulting rules was done by clicking a rules-tab. There were 2 versions:

Externalized: Arrows are only coloured (thus clickable) when an action is legal, and greyed out (unclickable) if a move is illegal. E.g. moving the dish empty from left to right in fig.3 is illegal because it violates rule no.3. This is externalized by greying out the pink arrows.

Internalized: All arrows are always coloured (clickable), providing no information about the legality of moves (allowing illegal moves). One can push all buttons at all times. If one wants to move the dish empty (illegal) by clicking the right pink arrow, the following happens: the dish moves to the right, but an error notification pops up saying "this is not possible". Subjects had no other choice than click "ok", and the move was undone.

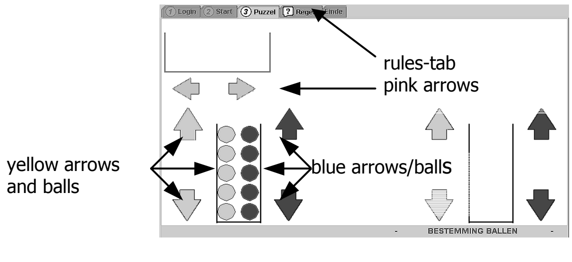

Figure 3 - The balls & boxes puzzle

Material Session 2

8 months later, in session 2, the same puzzle was used, but all subjects received an internalization-version of 5 trials. No rules were consultable, since we wanted to see

which knowledge was left in memory. Also a second puzzle was designed, “Wolves and Sheep” (W&S, fig.4) with the same problem space, rules and solution as B&B. This is an example of extremely near transfer, the only differences are the way the interface worked, the horizontal ordering of creatures, and more importantly their semantics (wolves eat sheep, boats cross rivers).

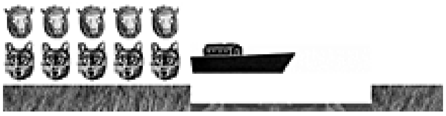

Figure 4 - The Wolves and Sheep puzzle

The third puzzle in session 2 was the classic Missionaries & Cannibals. It is similar, (fig. 5, only 3 M&C instead of 5 B&B) and semantically richer. After an error, the cannibals “really” ate missionaries (animation with sound). Here one has to apply (transfer) learned knowledge to a similar problem, and get used to the different amounts and reversed playing direction.

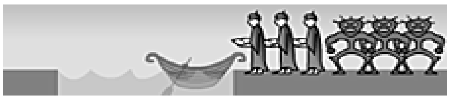

Figure 5 - The Missionaries & Cannibals puzzle

Procedure

In session 1 thirty subjects experienced with PC’s participated. The experiment had 3 phases (of 3 trials each), and a 10-minute interrupting distraction task between phases 2 and 3. The maximum time for each trial was 7 minutes. Different start-situations were used to avoid subjects relying too much on “having learned the trick” and repeat actions. In the second phase the play-direction of the puzzle was reversed. After the last trial, subjects filled out a knowledge test (score range 0-8) consisting of 4 MC and 4 open questions with screenshots of puzzle situations, they had to judge whether certain actions were possible. They also rated the clarity of the rules to solve the problem. Session 2 took place 8 months later. 14 subjects from session 1 (7 from each version) solved B&B again (5 trials) without further instructions, and filled out a knowledge test. Then, subjects had to solve W&S, followed by M&C.

RESULTS

We analysed the number of solved trials, time and steps needed, and how often dead-end-states were reached. The latter are states far from the solution with only one thing left to do: go back and find the right track. Although in the expected direction, differences in time and steps were not significant (not reported on here). Concerning dead-end-states, a MANOVA showed a nearly significant version effect (fig.6, table.1). Externalization-subjects reached more dead-end states, $F(1,28) = 3.58$; $p < .07$). There is also an interaction in phase 3 after interruption, internalization-subjects kept on improving, but the externalization-subjects performed worse than before ($F(2,56) = 2.11$; $p < .13$).

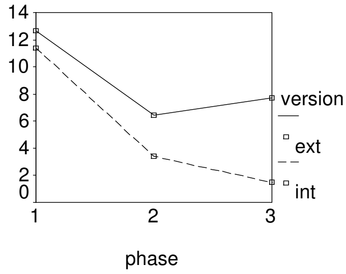

Figure 6 – Number of dead-end states per phase

		N	Mean	Sd.
	int	15	11.40	4.70
	ext	15	12.67	6.91
	int	15	3.40	3.18
	ext	15	6.47	7.11
	int	15	1.47	2.07
	ext	15	7.73	10.64

Table 1 – Nr. of dead end states per phase

For knowledge we found a main effect. The knowledge test (fig.7) showed that internalization-subjects had better knowledge than externalization-subjects (mean **7.33**, Sd 1.00 and mean **6.40**, Sd 1.24 respectively. T (28)=2.29, $p < .05$). There was also a trend on rating the question “the rules were clear enough to solve the problem” (fig.8, score range 1-5). Internalization-subjects found the rules clearer than the externalization-subjects (mean **4.13**, Sd 0.52 and mean **3.53**, Sd 1.25). T (df=28)=1.72, $p < .10$.

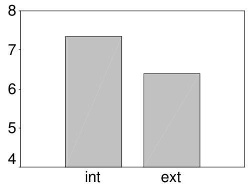

Figure 7 – Knowledge of the rules per version

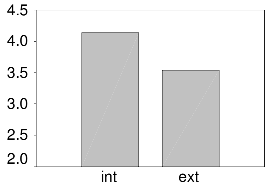

Figure 8 – Clarity of the rules per version

In session 2 converging results were found. In the first puzzle (fig.9, B&B) internalization-subjects needed significantly less time to solve it the first time (mean **432** sec, Sd 314 and mean **778** sec, Sd 397). T (df=12)=-1.81, $p < .05$. After this first success, subjects made few mistakes; all solved the puzzle equally well. Also the second task (W&S) was solved equally well.

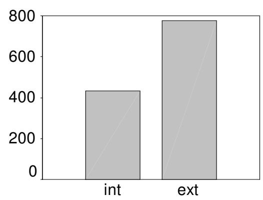

Figure 9 – Time (sec) until success B&B per version

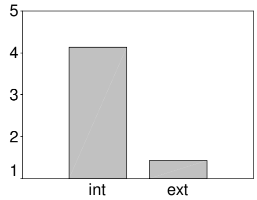

Figure 10 – Number of M&C solved per version

Looking at transfer however, internalization-subjects managed to solve the M&C (fig.10) 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 hypotheses stating that initially externalization yields better performance, was hardly supported. There were some differences concerning time and steps needed in the predicted direction, but not significantly so. The second hypothesis stating that internalization yields better performance in later stages was partly supported. For time and steps needed we found some tendencies in the expected direction, but no significant differences. However a more delicate measure "reaching dead-end states" informing us on how subjects behaved, in terms of the insight they had, we found an interesting result. We assumed that internalization-subjects do some smarter, more elaborated planning. Externalization-subjects were expected to solve more by trial and error and on base of interface cues, and therefore reach those problem-states more often, especially after a WM-erasing interruption. This was indeed the case. It showed that internalization-subjects performed better overall, reaching those dead-end problem-states less often. Furthermore there was the trend-like interaction that after an interruption, internalization-subjects kept improving while externalization-subjects fell back, depending on the interface, reaching more dead-end-states than before. This confirms our expectation that after an interruption, internalization-subjects continue to work on base of a plan-based strategy. The third hypothesis was supported; we expected internalization-subjects to have better knowledge after the experiment, because they had to build a stronger, more elaborated plan that relied less on interface information. Internalization-subjects had indeed better knowledge of the problem's rules and problem space. There was also a trend that internalization-subjects rated the clarity of the rules higher. This is intriguing, because externalization-subjects had interface feedback *and* consulted rules. Internalization-subjects, who *only* had the rules and no interface-help, found the rules clearer. We interpret the above as indicators of better understanding instigated by internalization. In session 2 internalization-subjects still had advantages. After not seeing the problem 8 months, they solved it several times again, and the first success took them only half the of time externalization-subjects needed. After this first success, all subjects remembered how to do it and performed equally well. A second puzzle, W&S with a slightly different look and feel yielded no differences. Yet another puzzle however, the classic M&C which was much more different, was solved almost 3 times more often in the same time by internalization-subjects. We interpret this in a transfer-context; internalization-subjects still have better imprinted knowledge, and facing a "new" situation, they have knowledge more readily available.

Some remarks remain. Perhaps in session 1, internalization-subjects might have done better if their application hadn't forced unintentional delays on them. They had to click away application-messages, and the program undid "wrong" moves. Because of this delay, subjects perhaps lost time in recovering, described by O'Hara and Payne (1999) as the effect of "lockout time and error recovery cost".

These experiments are a pilot study with rather abstract material to figure out if there is something to our notion that externalization might aid interaction, but that it does not necessarily aid understanding. The assistance introduced by providing context-sensitive feedback, proved to be of influence on acquired knowledge, transfer and performance on the same problem and a after a long delay.

One can look at an interface itself, at how well it facilitates learning a certain task such as "learning to use Photoshop". Here solid knowledge is not so crucial because making mistakes has no severe consequences. In other situations however it is extremely critical for an operator to know exactly what he is doing. Especially in an imaginary case where a system is down or fails and a process (e.g. shutting down a part of a factory) has to be performed without the system guiding you. In this case solid underlying knowledge of the task is highly necessary. From a slightly different angle, if knowledge acquisition ("learning something") is the aim, future research can also yield interesting results. Here we investigated the influence of internalized vs. externalized information on learning to solve a puzzle. If the results also prove to be true on other more realistic situations (think of learning a language with a CD-ROM, school situations where a physics principle is being thought) we can develop specific guidelines for specific tasks.

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. educational applications, spreadsheet applications or drawing applications with less repetitive and more complex tasks.

REFERENCES

1. Baddeley, A.D. and Hitch, G. (1974). "Working memory" . In G. Bower (Ed.), Recent advances in learning and motivation. New York, Academic Press.

2. 2. Nielsen, J. (1994). Usability Engineering. Morgan Kaufmann, San Francisco.

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

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

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

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

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