

# Externalization vs. Internalization: The influence on problem solving performance

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

## Abstract

*This research investigates how externalizing information in program-interfaces influences problem solving performance. Externalization as screen-representation provides relevant interface-feedback, and remembering certain task-related knowledge becomes unnecessary, relieving working memory (WM). Examples are visual feedback aids such as “graying out” non-applicable menu-items. Differently, when certain task-related information is unavailable, it needs to be internalized, stored in WM. We examined the hypotheses that externalization yields better performance during initial learning, while later, internalization yields better performance and knowledge. Two versions (internalized & externalized) of an isomorph of the “missionaries and cannibals” problem were tested. 30 subjects solved 6 problems, then performed a distraction task erasing WM, and solved 3 more problems. Also knowledge of the problem’s rules and states was tested. Internalization yielded better knowledge and reaching less dead-end problem-states. Time needed, steps and error-rates were in the expected direction, but not significantly. This research contributes to GUI design-guidelines for human-computer interaction of problem-solving tasks.*

## 1. Introduction

Software has seen a tremendous development during the last decades. Becoming more advanced, the amount, diversity and high turnover demand fast learning from users. Accents have been put on usability (effectiveness, efficiency & satisfaction) that users experience [2]; 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 [7]. Examples are wizards, help-options and graying-out menu-items that don’t permit using them thus offering a context-sensitive interface with just “possible” actions. For example in

Word, you cannot select “paste” from the “edit”-tab in the menu, when nothing is copied first (fig.1). The word “Paste” shown in gray color indicates its presence, but tells us that using it is impossible for some reason.

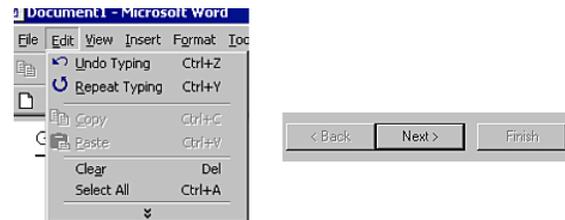


Figure 1 – Interface examples of graying-out

“Graying-out” is one example of **externalizing** information. By making information available on the interface, remembering certain information is not necessary. By externalizing information, working memory (WM) is relieved [9]. 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 [8, 9]. This was generally accepted, but as Zhang and Wang [9] 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 Working Memory Framework [1]. It has two components IWM (internal representations & memory processes) and EWM (external representations & perceptual processes). They showed that performing tasks with: *all* information in EWM was *always* easier than performing tasks with all the information (or partly) in IWM. Tasks with information distributed over IWM and EWM was either *easier*, or *more difficult*, dependent on how the information was present in IWM.

In computer programs, it is not always possible to have all cognitive functions externalized. We need a way to spread information over IWM and EWM effectively. To contribute to a more complete theory, research is

needed on whether other aspects of performance are affected by externalization. One could argue that a problem with full externalization is that users do not need to reason – they are not triggered to look for underlying rules, whilst they might be necessary in order to build up stable knowledge that also can be applied in new situations. The notion that learning is more effective when people experiment themselves is far from new, and exploratory learning has been a subject of research in many domains. More than a decade ago Carroll [3] already propagated minimalism in design and instructions. Perhaps having 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 tasks is important. O’Hara & Payne [4,5] and Trudel & Payne [6] 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 on the other hand makes little use of learned knowledge but relies on interface information. Plan-based activity leads to a shorter solution route, because steps are planned, and no unnecessary steps are taken, while a display-based strategy involves more steps because of more searching. Perhaps plan-based behavior results in still other favorable outcomes related to problem solving. A measure to look at could be the reaching of “dead-end states” in a problem space 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 program entailing the testing of our hypotheses. 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, and memory for the task after delay. In later phases we will test our hypotheses on situated tasks.

## 2. Hypotheses

We tried to investigate our questions using following hypotheses (fig. 1):

1. **Externalization will initially yield better task performance than internalization.** Internal knowledge-elements are not yet acquired, externalization will be of help
2. **Internalization yields better task performance later after a distraction task.** Internalization already results in better knowledge, leading to better performance.
3. **Internalization yields better knowledge.**

Not relying on externalized information makes a user plan more steps himself. Knowledge of rules will be better here. An experiment was conducted in which subjects solved a puzzle on a PC 9 times in 2 conditions: internalization and externalization. We expect the outcomes to take the following course:

**Phase 1:** 3 trials. Subjects don’t have needed knowledge available yet, externalization offers more interface cues, and performs better.

**Phase 2:** 3 trials. Performance in the 2 conditions will be equal, because internalization-subjects acquired internal knowledge.

**Phase 3:** 3 trials, after an interrupting distraction task (erasing WM). Internalization-subjects perform better because of better internalized knowledge and a more elaborated plan.

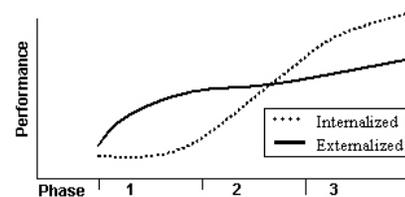


Figure 2 – Hypotheses

## 3. Method

### 3.1 Material

The problem-solving task we used is called “Balls & Boxes” (B&B) which is an isomorphic version of “Missionaries and Cannibals” (M&C). 5 missionaries and 5 cannibals are on a riverbank, and all have to reach the other bank by boat. Constraints are that the boat only fits 3 people, the minimum to sail is 1, and cannibals can never outnumber missionaries at any place, or the missionaries will be eaten. Our B&B problem (fig. 3) uses exactly the same problem space, but is more abstract. This was done because in M&C rules are quite easy to learn (Cannibals eat people, boats cross rivers etc.). Using boxes, blue and yellow balls and a dish instead, we avoided too easily learned rules. Rules were 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 boxes

The puzzle-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). Clicking on a rules-tab did consulting rules. The puzzle was designed in 2 versions:

**Externalized:** The arrows are colored (thus clickable) when an action is legal, and grayed out (unclickable, as in example fig.1) if a move is illegal. E.g. moving the dish empty from left to right (fig.2) is illegal, violating rule no.3. This is externalized by graying the pink arrows out.

**Internalized:** All arrows are always colored and clickable, providing no information about the legality moves (making illegal moves is possible). Subjects can push all buttons at all times. If one would want to transport the dish empty from left to right (illegal), the following would happen: the right pink-arrow is clicked and the empty dish moves to the right. The dish arrives right, but an error notification pops up saying “this is not possible”. Subjects had no other choice than click “ok” in the dialogue-box, and the move was undone.

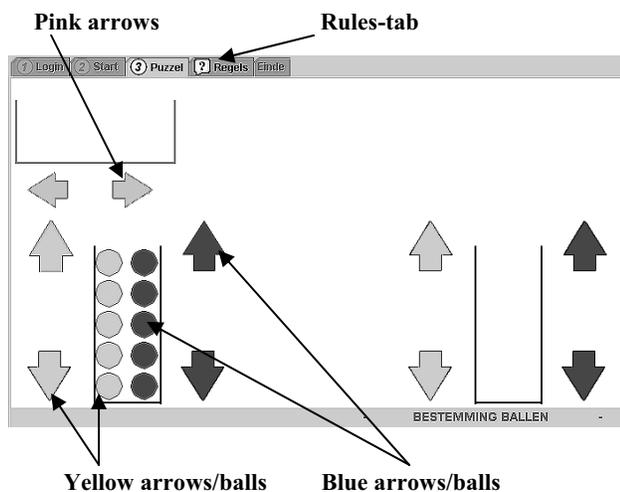


Figure 3 - The balls & boxes puzzle

### 3.2. Procedure

Thirty subjects of age 19-28, experienced with PC's, were recruited at Utrecht University. The puzzle, a Java-applet, ran on a Pentium IV 2,0 GHz PC with a 17" monitor. The course of the experiment was told, followed by a brief oral explanation of the interface and a short demonstration. The experiment consisted of 3 phases (with 3 trials each), and a 10-minute interrupting distraction task between phase 2 and 3. The maximum time for each trial was 7 minutes. Different start-situations were used to avoid that subjects relied too much on “having learned” the trick and simply repeat actions. Also, in the second phase, the play-direction of the puzzle was reversed from left-right to right-left. After the last trial, subjects filled out a knowledge test (score 1-8) consisting of 4 MC and 4 open questions with

screenshots of puzzle situations. They had to judge and explain whether, and why certain actions were possible. Subjects also rated how clear the rules were to solve the problem. After completing the experiment subjects received 5€ reward.

## 4. Results

We looked at the number of trials solved, time needed, the number of steps needed to solve them, and at how often dead-end states were reached. Dead-end-states are states far from the solution with only one thing left to do: go back and find the right track. In general, all subjects improved over time in both versions. Several measures were in the expected direction, but there were no significant effects, nor interaction effects of time and steps needed.

**Dead-end-states.** Below are the number of times per phase subjects reached dead-end states in the problem space.

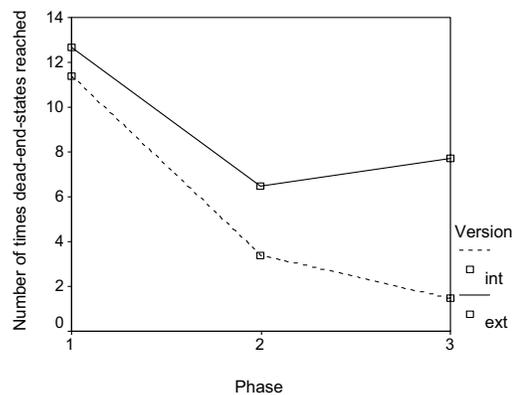


Figure 4 – Dead-end states per phase

		N	Average	Sd.
phase 1	int	15	11.40	4.70
	ext	15	12.67	6.91
phase 2	int	15	3.40	3.18
	ext	15	6.47	7.11
phase 3	int	15	1.47	2.07
	ext	15	7.73	10.64

Table 1 – Dead-end-states per phase

MANOVA showed an interesting nearly significant effect of version (fig.4, Table 1). Overall, externalization-subjects reached more dead-end states  $F(1,28) = 3.58$ ;  $p < .07$ ). In addition, there is a trend for an interaction ( $F(2,56) = 2.11$ ;  $p < .13$ ). In phase 3 after the interruption, in the internalized version performance kept improving,

but in the externalized version on the other hand, subjects performed worse than in phase 2.

**Knowledge.** Before measuring knowledge of the rules and states we confirmed that there was no difference in the amount of times subjects consulted rules. The knowledge test (scores 0-8) subjects received after the experiment, showed that subjects in the internalization-condition had better knowledge than subjects in the externalized condition (mean 7.33, sd 1.00 and mean 6.40, sd 1.24 respectively.  $T(28)=2.29, p < .05$ ).

**Estimated clarity of the rules.** Subjects also rated the question “the rules were clear enough to solve the problem”. There was the tendency that internalization-subjects found the rules clearer to solve the problem, than the subjects from the externalization condition (mean 4.13, sd 0.52 and mean 3.53, sd 1.25).  $T(df=28)=1.72, p < .10$ .

## 5. Discussion

Our first hypotheses stating that initially externalization yields better performance, was hardly supported. The differences between versions concerning time and steps needed, were in the predicted direction, but not significantly so. In the beginning internalization-subjects needed some more steps than externalization, which seems to point at a slight advantage of visual rule-feedback externalization enjoyed in the beginning, but later this difference disappeared.

The second hypotheses stating internalization yields better performance in later a stage was partly supported. For time and steps needed we found some tendencies in the expected direction, but no evidence. On another, more delicate measure however, we found an interesting result. We introduced this measure “reaching dead-end states” to inform us on how subjects behaved, in terms of the insight they had. We assumed that internalization-subjects do some smarter, more elaborated planning while externalization- subjects are expected to solve more by trial and error and on base of interface cues. Exactly because of this supposed difference and planning behavior we expected that externalization- subjects would get far more to those problem-states, especially after an interruption meant to erase Working Memory. This was indeed the case. It showed that internalization-subjects performed better; they reached those problem-states less. Furthermore, there was also the trend-like interaction that after the interruption, internalization-subjects kept improving, while externalization fell back, reaching more dead-end-states than they did before. This confirms our expectation that after an interruption, internalization-subjects continue to work on base of the plan-based

strategy as they did before, while externalization on the other hand performs worse after interruption. They fell back depending on the interface, having a less elaborated plan.

The third hypothesis in which we expected that internalization would remember more knowledge elements was supported. We assumed that internalization-subjects had to build a stronger, more elaborated plan and could rely less on interface information, and indeed working with the internalized version resulted in having significantly better knowledge of the problems rules and problem space. 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* consulted the rules. Internalization-subjects, who *only* had the rules and no interface help found the rules clearer. We carefully interpret these 2 findings as indicators of better understanding in the internalized condition.

Some remarks remain. Some things could have unintentionally influenced the results. Measures could have been different if subjects would have had more time. Also the 2 versions differed slightly; there were application-messages that had to be clicked away in the internalized version, and the program undid “wrong” moves. A small unintentional delay occurred, and maybe subjects lost time in recovering, described by O’Hara and Payne [5] as the effect of “lockout time and error recovery cost”.

We are curious to see how stable the better knowledge provoked by internalization is, and will continue the research and conduct the experiment again with the same subjects, after a delay of several months. We expect internalization-subjects to perform better because of their better knowledge. This isomorphic presentation of B&B was semantically less rich than the original M&C with fewer intuitive semantics to “colored balls” than to “missionaries and cannibals”. In the context of transfer we will investigate how subjects that worked with the 2 versions perform on yet another isomorph of “missionaries and cannibals”. To let outcomes of future research contribute to GUI design guidelines, we will further investigate different types of more realistic planning-related tasks, e.g. spreadsheet applications or drawing applications where actions are less repetitive and more complex.

## References

- [1] Baddeley, A.D. and Hitch, G., “Working memory”. In G. Bower (Ed.), *Recent advances in learning and motivation*. New York, Academic Press, 1974.

- [2] Nielsen, J. *Usability Engineering*, Morgan Kaufmann, San Francisco, 1994.
- [3] Carrol, J.M. (1990). *The Nurnberg Funnel: Designing Minimalist Instruction for Practical Computer Skill* Cambridge, Mass: MIT Press.
- [4] K. O'Hara and S.J. Payne, "The effects of operator implementation cost on planfulness of problem solving and learning", *Cognitive Psychology*, 35, 1998, 34-70.
- [5] K. O'Hara and S.J. Payne, "Planning and the user interface: the effects of lockout time and error recovery cost", *International Journal of Human-Computer Studies*, 50, 41-49, 1999.
- [6] C.I. Trudel and S.J. Payne, "Self-monitoring during exploration of an interactive device", *International Journal of Human-Computer Studies*, 45, 1996, 723-747.
- [7] H. Van Oostendorp and S. de Mul, "Learning by exploration: Thinking aloud while exploring an information system", *Instructional Science*, 27, 1999, 269-284.
- [8] J. Zhang, "The nature of external representations in problem solving", *Cognitive Science*, 21(2), 1997, 179-217.
- [9] J. Zhang and H. Wang, "An Exploration of the Relations Between External Representations and Working Memory", 1998, (in review, manuscript submitted for publication).