Practice Effects on Interruption Tolerance in Algebraic Problem-Solving

Diana Woelki (diana.woelki@gmail.com)
Berlin Institute of Technology, Department of Psychology,
Straße des 17. Juni 135, 10623 Berlin, Germany

Antti Oulasvirta (oulasvir@hiit.fi)
Helsinki Institute for Information Technology HIIT,
P.O. Box 9800, FIN-02015 TKK, Finland

Jürgen Kiefer (jki@ zmms.tu-berlin.de)
Berlin Institute of Technology (Center of Human-Machine Systems - prometei)
Franklinstr. 28-29, 10589 Berlin, Germany

Robert Lischke (lischker@student.hu-berlin.de)
Humboldt-Universität zu Berlin (Institut für Psychologie)
Berlin, Germany

Abstract

In this study, we examine the hypothesis of Ericsson and Kintsch's (1995) Long-Term Working Memory (LTWM) theory, according to which, with practice, people can utilize long-term memory so efficiently that they can overcome interruptions virtually without any costs. Six subjects were recruited to perform algebraic problem solving tasks for a total of nine hours in three consecutive days. Color patch n-back tasks interrupted performance, after which there was a recall task for the previous equation. Subjects' performance in the primary task increased across the three days, reflecting a general learning effect. More importantly, the negative effects of interruptions on memory recall decreased faster than the general learning effect predicted. However, this effect was small and limited to the two first days of the experiment. We discuss alternative explanations to this short-lived effect of practice on interruption tolerance.

Keywords: Interruption tolerance; memory skills; long-term working memory; retrieval structures.

Introduction

How does practice affect one's susceptibility to interruptions' disruptive effects? In this paper, we want to examine memory in relation to the complex domain of mental arithmetic. Ericsson and Kintsch's (1995) theory on long-term working memory (LTWM) argues that models of short-term memory are not able to explain experts' abilities to overcome disruptive effects. They propose that the best way to disentangle the division of information between the short- and long-term components is to examine "performance under unusual circumstances, such as interruptions imposed by switching between different tasks, by memory testing during processing, and by memory performance after processing has completed" (p. 240).

According to their theory, selective and anticipatory encoding into LTWM ensures direct access to long-term memory and takes precautions to prevent loss. The key postulate in explaining how LTM is utilized is called *retrieval structure*. Retrieval structures are abstract, hierarchical knowledge representations that experts develop to efficiently encode and retrieve information in LTM. Retrieval structures can consist of two types of associations. First, associations to a system of cues are repeatedly extracted to index specific semantic categories of information, the function of which is to "allow retrieval of the most recently encoded information through reinstatement of those retrieval cues, even when this type of information is frequently updated and changed during the processing" (Ericsson & Delaney, 1999, p. 579). Second, these cues can be embedded in generated structures in LTM, where presented information is interrelated with other pieces of presented or generated information. Oulasvirta and Saariluoma (2006) observed that subjects could reliably reinstate retrieval structures after a demanding interruption. They argued that once information is encoded into LTWM, content-based cues can be used to re-access the representations upon task resumption.

According to Baddeley's theory of short-term working memory (STWM), mental arithmetic relies on the phonological loop for temporary storage, whereas the central executive helps accessing and executing appropriate algorithms (Baddeley & Logie, 1999). STWM storage implies a fixed capacity, so that susceptibility to disruptive effects would not be expected to change with practice, given that storage to LTM has *not* taken place and that the interruption is long enough and so demanding that it prevents active maintenance of contents in STWM.

Altmann and Trafton (2003) claim that preparatory perceptual and memory processes can take place, at times spontaneously, and can mitigate the disruptive effects of task interruption. Rehearsal is one such process. They also hold the non-availability of cues responsible for loss of memory. This suggests that there are two conditions when interruptions should be particularly disruptive: 1) when subjects do not have time to engage in preparatory processes, i.e. when the interruption is immediate and there is no "interruption lag" between the main task and the interruption during which one could start rehearsal, and 2) when the recall task does not involve any cues that could help in retrieval. In our experiment, we have arranged that there is no interruption lag and the memory test is free recall.

One hypothesis is that practice is linked to strengthened knowledge associations. Siegler and Shipley's (1995) Adaptive Strategy Choice Model (ASCM) involves a hypothesis that with practice, more problem–strategy associations are established. Less efficient *procedural* strategies should be replaced by more efficient ones, and novel *retrieval* strategies should be adopted. But as demonstrated for mathematical skill, usage of both strategies depends on association strengths (Imbo & Vandierendonck, submitted).

Though there is more and more understanding about skills necessary for maintaining information during interruptions and re-accessing them (see Oulasvirta, 2006), the effects of practice on these skills are poorly understood. In this paper, we are interested in abilities achieved with practice that help to overcome interruptions' disruptive effects. The question of effects of practice is of importance for everyday settings where interruptions occur. If practice helps overcoming effects of interruptions, maybe this effect can be further reinforced by means of organization of work, user interface design, or training.

We examine algebraic problem-solving as the task domain. With an increasing amount of practice, subjects should improve performance on the primary task (here: solving equations for *x*). Because in this paradigm subjects have to solve the tasks mentally without external aids, skills should emerge for storing intermediate results necessary to solve the task. The gradually acquired skills should enable subjects to anticipate future retrieval demands, store necessary cues to these structures, and help them resume primary task representation. The amount of practice spent on the primary task should be positively associated with a decrease in interruption costs over time.

Method

We chose mental algebra as the experimental task, because mental calculation has high demands on working-memory and is well established in research of expertise (Ericsson & Kintsch, 1995).

To analyze effects of practice, subjects' memory was tested in terms of free recall of the previously seen equation. The key dependent variable is the accuracy recall that we measure after interrupted and non-interrupted processing. To construct more sophisticated measures of accuracy, every input string was parsed and compared along selected features of the presented equations. Position and occurrence of all elements within the equation were consolidated into separate dimensions such as structure, intermediate results, correct numbers, number next to the unknown variable *x,* and the Levenshtein distance (Levenshtein, 1966), a metric string for measuring edit distance. To analyze the time needed for an answer, a further calculation measured the reaction time needed to complete entering the equation and then divided it by each character inserted into the blank field. We expected to see a practice effect after roughly nine hours of algebraic problem-solving. In total, a subject practiced about 450 equations.

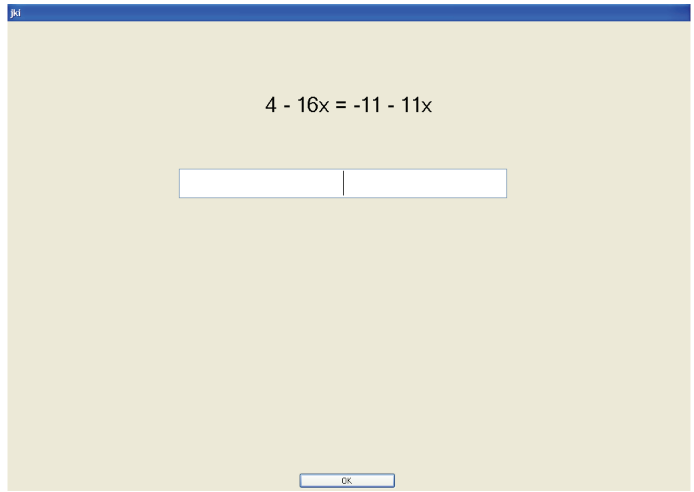

Figure 1: The experimental task is to fill in the x that solves the equation shown above.

Participants

Six undergraduate students (19-23 years; M=21.5; SD=1.5) from Finland and Germany, three females and three males, were recruited for the study. They attested to having average school knowledge in algebra and not to study mathematical sciences.

Materials

The experiment was divided into six single sessions, these in turn were divided into three days. To make sessions comparable, all sessions involved the same formal set-up in regards to the content. No equation was shown more than once per subject.

In each session, subjects were asked to solve 75 equations. The difficulty of the equations was randomized in terms of complexity of strategies needed to solve *x*. In the first level, equations were quite easy to solve (e.g. x+3=16-5), whereas the last level took much more effort (e.g. –91x+36=-80(x-1)). Equations varied in calculation type (addition, subtraction, multiplication, division), value of numbers and use of parentheses. No external aids, such as paper and pencil, were permitted. Each equation was always on the screen when solving it until "OK" was pressed (see Figure 1).

In order to create an engaging interruption task, we chose the *n-back task*, which is used in a wide range of neuropsychological, clinical and cognitive studies and tests. When tested upon the cortical regions involved, it proved to equally demand working memory, no matter if stimuli were verbal or non-verbal (Owen et al., 2005). Instead of numbers or letters, we chose *a colored n-back task* (Gevins &

Cutillo, 1993) with unicolored screens to prevent interference between primary and n-back task. According to Edwards and Gronlund (1998), the less similar the materials between the main and the interruption task, the less memory loss is to be expected from interference. Moreover, we adapted the n parameter to subjects' maximum that we estimated individually in a pre-experimental test. The difficulty of the n-back task varied from two to four items to be remembered backwards. When the task itself was actuated, it displayed sequences of six colors (red, green, blue, yellow, grey and black) one at a time. Each color patch appeared for *three seconds* in a randomized fashion one after another on a full screen. The task required subjects to match every color with the color n times before, by pressing labeled "yes" and "no" buttons on the keyboard. If an answer was skipped, it counted as a miss. A ratio of 1:3 yes vs. no answers was enforced by the program, in order to keep subjects concentrated.

The memory test task required subjects to type the exact presentation of the previous equation into a blank field. In the interrupted condition, the memory task appeared right after the interruption. In the continuous, non-interrupted condition it popped up as soon as the equation had been solved.

Feedback was provided after each session. The program informed the subjects about the percentage of their correct answers given in the primary and n-back task, but not for the memory task.

Procedure

When a result for *x* was entered into a blank field and "OK" was pressed, the next equation appeared. Returning to a previous task was prevented. The n-back task appeared unpredictably and twice per difficulty level. Before the program switched to the n-back task, subjects were given ten seconds to think about the equation. They were then required to pay attention to the n-back task for *one full minute*.

The memory test followed all interruptions. It also appeared just as often *without* an interruption (the non-interrupted condition). In this manner, each subject was tested on an equal number of recall conditions. The interruption effect could be constructed as a dependent variable by comparing interrupted to continuous performance.

The sessions were equally spread on three consecutive days, so that each subject had to work on two sessions per day. All six sessions were followed by two additional equations not included in the data analysis. Here, subjects were instructed to give think aloud while solving the two tasks. They were trained for this before the actual experiment. At the end of the whole experiment, subjects were asked in a short interview how they had tried to remember the last equation.

Subjects could always choose autonomously when to start solving an equation, though they were still instructed to do as well and quickly as they possibly could. N-back tasks were experimenter-paced and were to be done as accurately as possible. On average, one session lasted for 90 minutes.

Results

To go through the results, we start with general learning effects in the main task and then focus on interruption costs. The results show a general practice effect in the primary task and an effect of practice on interruption tolerance that is relatively small and limited to the two first days of the experiment.

Primary Task

To understand effects of practice on the primary task, and so to avoid confusing effects on interruption tolerance to general learning effect, we ran a paired samples t-test to compare subjects' mean primary task performance of the first three sessions (M=10.2, SD=0.96) with those of the last three sessions (M=11.4, SD=1.51). The test yielded an expected increase from the first half to the latter one in the variable *correctly solved equations,* $t(5)=-3.8$, $p<0.013$ (see Figure 2). At the same time, reaction times dropped from the first three sessions (M=44.97, SD=9.25) to the last three sessions (M= 36.7, SD=6.3), $t(5)= 2.7$, $p< 0.031$, (see Figure 3).

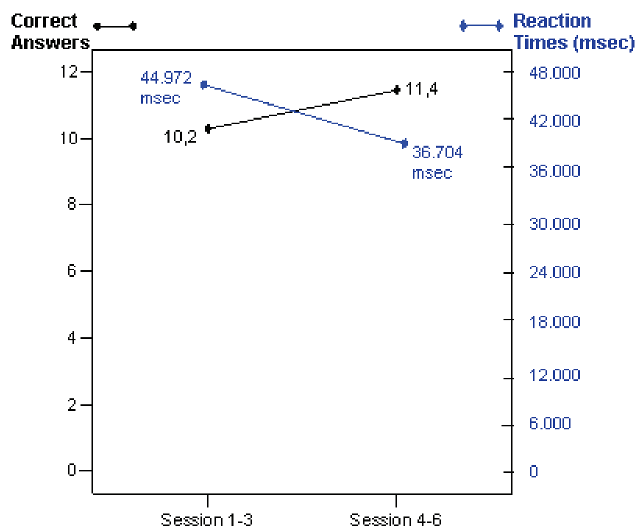

Figure 2: Mean number of correctly solved equations and mean reaction time per equation (msec.) of the first three vs. last three sessions. Shows a general practice effect.

These findings echo with what we found in the verbal protocols. When thinking out loud, subjects' mental processing became more organized while solutions were found with less effort. To sum up, these results show that the subjects improved in their ability to solve algebraic tasks.

General Interruption Costs

We next turned to *memory performance*, comparing interrupted to non-interrupted condition. One subject was excluded from the analysis for this variable, because of lack of patience on the third day in trying to remember the equations, which led him/her to insert only the answers for *x*.

The measure of *correct numbers* was the only one showing effects. All other t-test values for interrupted vs. non-interrupted were below the threshold (e.g. for structure $t(4)$=-0.198, p< 0.85; Levenshtein $t(4)$=-1.2, p< 0.29). In this and following sections, "memory performance" refers to the variable *correct numbers* in free recall. First and foremost, overall memory performance in the continuous task (M=0.57, SD=0.21) was almost three times as good as performance in the interrupted task (M=0.18, SD=0.06). This manifests a disruptive main effect of interruptions.

Practice Effects and Interruption Costs

We then wanted to examine if practice affected interruption tolerance. We therefore ran multiple paired-sample t-tests to find out about performance changes *within each day*. First, we separately examined the changes within interrupted and continuous memory performance by comparing *correct numbers* of each day's first session (sessions one, three, and five) to the corresponding second sessions (sessions two, four, and six).

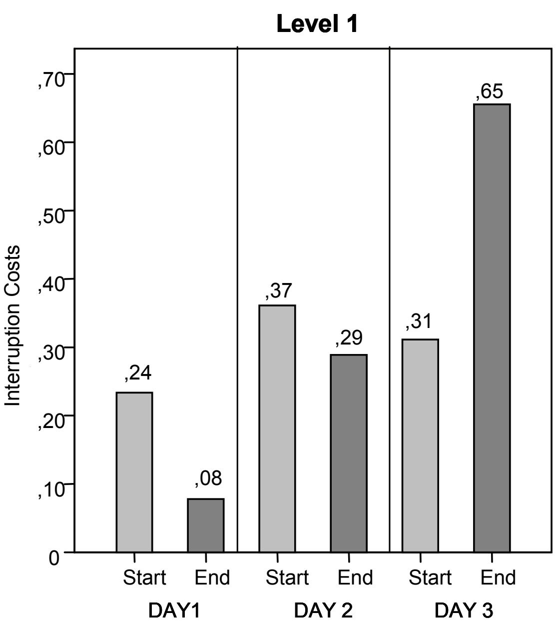

Figure 3: Interruption costs for each day's starting and ending session. To calculate interruption cost, scores in the interrupted condition were subtracted from those of the continuous condition. This graph includes data from the easiest equations (difficulty level 1).

In Figure 3, costs of interruption are calculated by subtracting interrupted from continuous recall performance *for the first (easiest) difficulty level*. Whereas costs decrease in the course of day one and two, they rise up on the third day. It may be that the subjects felt underchallenged by these easier equations after practicing hundreds of more difficult ones.

Because overall performance of all subjects was much weaker on the third day, we subsequently focused on the first two days.

Memory performance in the continuous condition did not improve as much in the first two days, combined, as in the interrupted condition, $t(4)$=-0.310, p<0.77. Memory performance in the interrupted condition improved significantly better than in the continous condition within a day (from M=0.15, SD=0.04 to M=0.26, SD=0,06), $t(4)$=-5.401, p<0.006 (Figure 4).

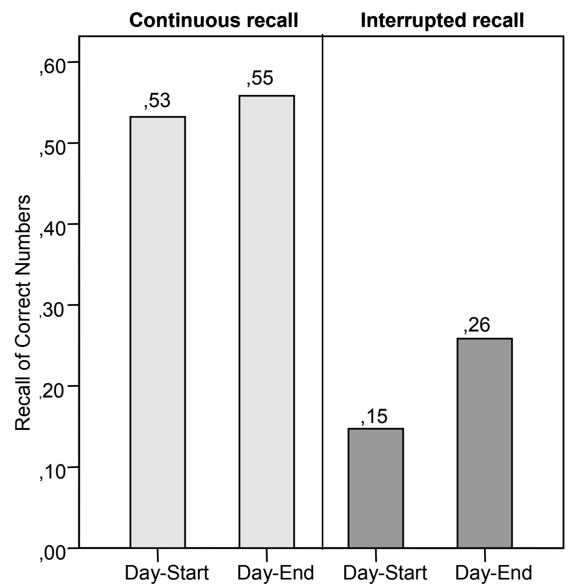

Figure 4: Memory performance of the first two days combined, divided into the starting and ending sessions completed, for continuous recall (p< 0.77) and interrupted recall (p< 0.006).

Figure 3 examined the difficulty level 1 only, and we wanted to see how if the pattern observed there holds for all difficulty levels. Consequently, memory performance of a day's first session was subtracted from the last session of that day. Within-day improvement was thus divided into the three days. Looking at Figure 5, improvement in interruption tolerance within a day, as measured by comparing performance in the day's last and first sessions, is present only for the two first days of the experiment. A cross-comparison with Figure 3 hints that the leveling off may be due to difficulty level 1.

Do these differences between continuous and interrupted recall represent a more robust improvement in interruption tolerance that behaves differently from the general learning effect? With respect to *all three days*, overall performance improvement in continuous recall (M=0.007, SD=0.047) was compared to improvement in interrupted recall (M=0.103, SD=0.025) yielding a rise in interruption tolerance, ($t(2)$=-7.25), p<0.018.

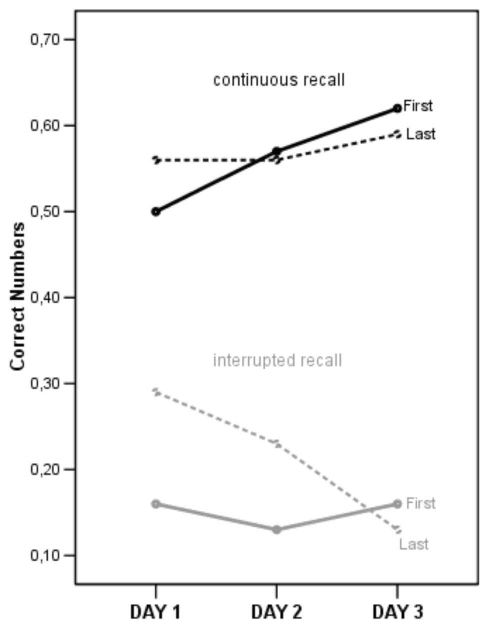

Figure 5: Within-day performance differences for continuous and interrupted memory as measured by correctly remembered numbers of the equation.

We finally examined reaction times in the recall task. We ran a paired-samples t-test and compared the reaction times of the interrupted memory task of the first three sessions (M=6.6 s, SD=3.3) to the times of the last three sessions (M=4.1 s, SD=2.5). Results show that the time needed to retrieve the previous task representation decreased significantly, $t(4)=1.173$, $p<0.045$, too (see Figure 6). Hence, while subjects improved in interrupted memory performance, they needed even less time.

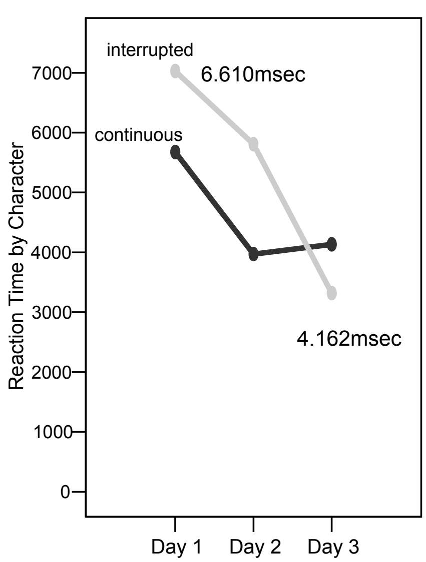

Figure 6: Mean reaction time divided by character for the continuous ($p<0.31$) and interrupted memory task ($p<0.045$)

Discussion

Subjects' performance in the primary task increased across the three days, reflecting a general practice effect. More importantly, the negative effects of interruptions on memory recall decreased faster than the general learning effect predicted. However, the effect of practice on interruption tolerance was limited to the first and second day.

The subjects stated in the interviews that rehearsal of the equation was almost impossible during the n-back task, as it demanded too much of their attention. Practice effects can thus not be trivially attributed to increase in ability to utilize intermittent rehearsal. We hold it unlikely that rehearsal was used as a strategy to maintain task representations during interruptions. Therefore, it remains to be explained how LTM may have been employed here.

Our dependent measures provide some tentative evidence for this question. Of all the variables we used to measure memory accuracy (e.g., structure, intermediate results, edit distance and numbers next to *x*), the only one showing improvement with practice was the variable *correct numbers*. Verbal protocols hinted that subjects were "reverse engineering" the to-be-recalled equation from the solution they remembered. They commented on how they remembered exact numbers by "counting back" from end results.

Why was the interruption cost still very large? We expected that nine hours of practice would nearly eliminate interruption costs (see Edwards & Gronlund, 1998; Trafton et al., 2003). One explanation is that the experimental paradigm represents a sort of "rote practice" learning where "practice" means repeating the same task over and over again under time pressure. The conditions leave virtually no room for trying out alternative strategies. Nevertheless, we did observe a shift in strategies that may have helped the subjects in stabilizing their representation in LTM and thus recall more elements. In the first session the subjects started by moving constants to the right hand *side one at a time*, then summed them up, and finally summed *x*s on the other side. Toward the last session, the subjects seemed to move *all* constants to the right side *in one step* and sum them up. The transition to the move-all strategy did not happen at once but gradually. (Parenthetically, this gradual shift may help explaining part of the general learning effect.) The presence of a large interruption effect even in the very last session suggests that the subjects were still relying on STWM despite after many hours of experiment.

What explains, then, that the fact that the observed practice effect was limited within a day? Our hypothesis combines two assumptions: 1) more stable retrieval structures achievable due to the strategy change and 2) increase in the activation levels of seen patterns in LTM. When solving an equation, even though the equation is in front of you on the display, you are going further and further away from the given starting point with each step taken. Assuming that you get faster in solving reoccurring small-integer operations (e.g., "8*9"), perhaps due to increase in activation levels of their representations in LTM, it means that you have probably advanced *further* in solving the equation when the interruption occurs—which in this experiment occurred after a fixed interval. This in turn means that you have *less* objects to keep in your working memory at that point, as each step you have taken has decreased the number of elements that have to be kept in mind. Moreover, if one has advanced further in the solution tree, the last element one remembers after an interruption is going to be more informative of the original equation as the representation is less fragmented. By contrast, if one is interrupted while doing the very first steps, one has built no relationships in LTM between the items. Simple activation-based account can thus cover the within-day improvement we witnessed. This effect may be boosted by the strategy change we observed: in the beginning stages a cue that one remembers is *arbitrarily* related to the original equation, but later on one when the procedure stabilizes and each type of element has its own predictable position in the retrieval structure, one knows *to which stage of the solution tree* a given cue type (e.g., "4x") belongs to.

But why would these effects limited to a day, given that there was a global practice effect in the main task? Why would time reset the positive effect of practice on interruption tolerance? Some of the many factors that could have influenced the results as main effects or in interaction with the strategy changes involve time of day of testing, delay since last testing session, and interspersed sleep. Our current hypothesis is that activation built for frequently seen patterns (e.g., 7*7=49) levels off with passing time. Thus, after a night's sleep, one cannot as quickly "see" the solutions to items (e.g., 8*9), with the consequence that one is a bit slower in the beginning and, importantly, more likely to have advanced less before the interruption occurrs. Thus, one is again in the situation where the elements that can be recalled are less informative of the whole equation. Future studies will attempt to localize the factors responsible for these effects.

Interestingly, the effect of practice on interruption costs seemed to level off on the third day. Motivational reasons may be one reason for poorer performance on the last day. The experiment is gruelling. Alternatively, it may be that they reached a plateau in learning and need qualitative changes in their problem-solving strategies to develop more elaborate and comprehensive retrieval structures.

This paper provides an interesting message to applied cognitive scientists: positive effects of practice on interruption tolerance can emerge within a single day of practice. Many of the experiments in, for example, human-computer interaction, have been conducted in conditions where the subjects have been unfamiliar with the task and the experimental materials, thus creating conditions for strong interruption effects to appear. It is possible that, in those activities in which people have training and practice, they do not suffer that much from interruptions. It remains to be examined what kind of task or work conditions favor even stronger positive effects

Acknowledgments

We thank the Center of Human- Machine Systems of the Technical University of Berlin and its research training group *prometei* for providing resources and financial support to run the experiments. AO acknowledges support from the Emil Aaltonen Foundation. We thank K. Anders Ericsson and Pertti Saariluoma for important comments.

References

Altmann, E.M., & Trafton, J.G. (2004). Task Interruption: Resumption lag and the role of cues. *Proceedings of the 26th annual conference of the Cognitive Science Society*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Baddeley, A.D., & Logie, R.H. (1999). The Multiple-component model. In Miyake, A., Shah, P. (Eds.), *Models of Working Memory Mechanisms of Active Maintenance and Executive Control* (pp. 28-61). Cambridge, UK: Cambridge University Press.

Ericsson, K.A., & Delaney, P.F. (1999). Long-term working memory as an alternative to capacity models of working memory in everyday skilled performance. In Miyake, A Shah, P. (Eds.), *Models of Working Memory Mechanisms of Active Maintenance and Executive Control* (pp. 257-297). Cambridge, UK: Cambridge University Press.

Ericsson, K.A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review, 102.*

Gevins, A.S., & Cutillo, B.C. (1993). Neuroelectric evidence for distributed processing in human working memory. *Electrophysiology and clinical Neurophysiology, 87*, 128-138.

Imbo, I., & Vandierendonck, A. (submitted). Practice effects on strategy selection and strategy efficiency in simple mental arithmetic. A manuscript submitted for review.

Levenshtein, V.I. (1966). Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Physics Doklady, 10*, 707–710.

Oulasvirta, A., 2006. *Studies of working memory in interrupted human-computer interaction.* Doctoral dissertation, Department of Psychology, University of Helsinki.

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

Owen, A.M., McMillan, K.M., Laird, A.R., & Bullmore, E. (2005). N-back working memory paradigm: A meta-analysis of normative functional neuroimaging studies. *Human Brain Mapping, 25*, 46-59.

Siegler, R.S., & Shipley, C. (1995). Variation, selection and cognitive change. In Simon, T.J., Halford, G.S. (Eds.), *Developing Cognitive Competence: New Approaches to Process.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Trafton, J.G., Altmann, E.M., Brock, D.P., & Mintz, F.E. (2003). Preparing to resume an interrupted task: Effects of prospective goal encoding and retrospective rehearsal. *International Journal of Human-Computer Studies, 58*,583-603.