Psychology Bachelor thesis. Graduate theme: Technology and Media.

The effects of incidental background colour cueing on motor skill acquisition and transfer of training in a VR version of the Tower of London spatial problem solving Task

Christian Wanrooij
University of Twente

ABSTRACT

The aim of this study was to investigate whether motor skills acquired during training on a virtual Tower of London task, are in fact dependent on contextual factors of the learning interface. More specifically, whether background colour serves as a relevant incidental cue to encoding the skills involved in moving virtual rings through operating a mouse. The rationale behind it is that the more automated, the more context dependent a skill becomes. This holds that with practice, skill is increasingly more dependent on incidental cues, which (thus) allows for a reduction of cognitive load invested in the task. The main hypothesis stated that after sufficient training, and due to developing contextual dependencies, skilled participants would demonstrate a sudden decline in performance when transferring their skills to a Tower of London task where some context variable (i.e. background colour) differed from the learning environment. Results reflected the expected trend but did not reach statistical significance, though decisive conclusions can certainly not be drawn just yet. A critical review of the used software brought to light some flaws that might very well have influenced the experimental outcome. Some thoughts on why this version of Tower of London might not be a good tool for measuring the intended effect in this experimental setup are discussed.

1. Introduction

Virtual reality (VR) is truly an infant technology in the serious sense that its potential applicability is as tremendous as it is diverse in an equally diverse range of markets, its features, however, long from matured or even understood. The world of the virtual may take any shape or form, conform any purpose for which we lend it existence, though a decent understanding of the axis of its viability, solid psychological theory, is pending. So the question is: Building on psychological theory and models of information processing, how can we conceive the ultimate learning environment and guarantee optimum transfer of skill to real-life situations? The answer is not easy, because, as will become clear, skill acquisition is more than our conscious perception of it.

Computers have, at least in western societies, become almost as common a household appliance as the vacuum cleaner has been for many decades. Though computers are inherently complex devices, millions of people are able to operate them through a layer of software linking internal complexity to easily operable command prompts, windows, icons, menu's and pointers in the computer screen's interface. To optimize user friendliness, manipulation, manoeuvrability and navigation in the interface are – at least to a large degree - programmed to correspond to the physics of the real world. The point is that we are by no means strangers to VR. Computers are not just appliances in the classical sense. They aren't classical media either. We interact with and sometimes find ourselves meandering in or actively searching the virtual environments the computer grants us access to (e.g. the internet). Research has shown that people do not instinctively distinguish between what is real and what is only seemingly real – this is hardly surprising; we are, after all, not *homo technologicus* - which holds some dramatic implications. For instance, people tend to socially interact with computers (and other media) as if they were real people (Reeves & Nass, 1996), experience physiological changes in response to events in virtual environments (e.g. Jang et al, 2002), even so far that they can experience social anxiety in a virtual crowd (James, Lin, Steed & Slater, 2003). For us, apparently, a virtual environment does not differ much from the real world. In fact, virtual environments are now, and have been for a while, even used to *substitute* reality. In simulator training, a substitute for reality is created that has the look and feel of the real deal, but that allows you to make consequence free mistakes that might have been costly, even deadly, in real life. Simulator training has proved an efficient, cost effective, flexible and safe means by which to train people in acquiring a large variety of skills. Numerous industries such as the military, medicine, and traffic have all employed simulators as powerful training tools. Simulator assisted training significantly improves training efficiency in training an array of different skills (e.g. Hayes, Jacobs, Prince & Salas, 1992; Mani, Shoor & Pedersen, 2004; Peugnet, Dubois, & Rouland, 1998; White, Carson & Wilbourne, 1991). Indeed, simulator training improves not just performance, but knowledge, confidence and reaction as well (Doerr, Quinones, Dipboye, Dunbar, 2000).

But despite these obvious merits of simulator training or 'virtual learning', there is still need for further research into the psychology of dealing with virtual environments so that transfer of training to real-life situations can be optimized. Most simulators are built to produce high levels of realism in the hope of reaching substantial transfer of training. It is obvious that this endeavor is an expensive one. It might be a waste of monetary funds considering that it is not even clear to what degree realism equals optimal transfer. Multiple studies have, in fact, shown that high fidelity does not necessarily produce training advantages at all (Druckman & Bjork, 1994; see also Kozak et al., 1993; Lathan et al., 2002; Schneider, 1985). There are, however, some studies that seem to conclude somewhat to the contrary (e.g. White et al, 1991; see Barry, Mcgaghie, Petrusa, Lee & Scalese, 2005, for a literature review on high fidelity simulator effectiveness). It might even be possible that for some skills, high fidelity is necessary for adequate transfer of training, while for others it might not be at all. If different tasks would require different levels of fidelity for optimal skill acquisition, this would surely complicate matters. Research into the dynamics is necessary.

The costs of VR training can be trimmed down dramatically and its effectiveness increased if we reach a greater understanding of what exactly makes VR an effective training tool and can provide a sound psychological framework for virtual learning. Exactly what degree of realism is needed for optimizing training efficiency on a certain skill? Which aspects of the (virtual) environment contribute to acquiring and/or transferring skill? Most tasks learned in training simulators are very complex and involve many different skills, both motor and cognitive. To what degree do these different skills transfer to real life? Is there congruence or not? Most of these questions are beyond the scope of this article, but highlight the need for more research into this field.

1.2. Goals

In trying to reach a greater understanding of issues concerning VR fidelity and optimization of VR training efficiency, this article aims to function as the impulse for perhaps an agenda for research on exactly which contextual factors in VR are relevant, rather than irrelevant to skills acquisition and thus, transfer. The present experiment will investigate the role that background colour to a virtual Tower of London task (e.g. Shallice, 1982; Rowe, Owen, Johnsrude and Passingham, 2001) plays in encoding skill as well as the effect it might have on transfer of training. For research into such issues, there is a tremendous amount of psychological and interdisciplinary research literature to draw from, although scientific links to virtual learning, task-, skill- and contextual specifics still have to be made. Summing up, the hypothesis states that changing the background colour to a virtual Tower of London task after a certain time for training will negatively affect motor skill, due to developed contextual dependencies.

1.3. Theoretical justification

When people process information, this information is always embedded in the context within which it is presented. We certainly cannot disregard this context, for memory representation for items, events or procedures include additional information present at encoding, that may later serve as a prompt to aid retrieval of this information. (e.g., Anderson, Wright & Immink, 1998; Wright & Shea, 1991)

Over a century ago, Thorndike and Woodworth proposed the *identical elements theory of transfer* (Thorndike & Woodworth, 1901). The theory proposed that the amount of transfer training in one situation would have on another would be determined by the number of elements that the two situations had in common. Although their view now seems overly deterministic on the part of contextual elements, their idea would prove to have some merit. Thorndike and Woodworth’s ideas seemed in line with that of, in many respects, their successor, Pavlov. In correspondence with his connectionist approach and his discovery of the stimulus-response pairing in classical conditioning, Pavlov stated that the more similar the conditions of practice were to the criterion task, the greater the transfer (Pavlov, 1928). Theories describing context dependent learning came to be known as specificity theories and were the subject of many studies (see: Hull, 1943; Guthrie, 1950; Guthrie, 1952; Henry, 1958).

Following these studies, a sound theoretical framework for memory of ideas and events was provided in the *encoding specificity principle* of learning (Tulving and Osler, 1968; Tulving and Thomson, 1973; Hintzman, 1990). Its basic premise is that memory is improved when information available at encoding is also available at retrieval: The way in which memory of an item, idea or event is represented in the mind includes additional information present at encoding. If some part of the context is stored as a part of memory, it can serve as a cue for retrieving the remembered relevant information. One of the most dramatic illustrations of this principle was demonstrated in an experiment by Godden and Baddeley (1975), which showed that lists of unrelated words learned either in the water or on the deck of a ship were best remembered and retrieved under that same condition; on deck or in the water. Although the encoding specificity principle concerns only episodic memory, it would seem at least plausible that whether it concerns episodic or procedural memory, contextual information plays a vital part. Is it thinkable that the underlying processes of learning, whether from a cognitive or motor standpoint, are similar? If so, then the specificity principle might have great bearing to motor skill acquisition. Although more research into the specifics is needed, some findings are strongly pointing in this direction. Indeed, Ma, Trombly and Robinson-Podolski (1999) found a relation between context and motor skill acquisition and transfer. They tested whether participants would acquire and transfer their skills using chopsticks better in a natural environment than in a simulated setting. The natural context elicited significantly larger improvement of success rate in the acquisition phase and a significantly higher success rate in the transfer phase than the simulated context. These results suggested the use of (natural) contexts to facilitate the outcome of motor skill learning. Wright and Shea (1991) found that during motor skill acquisition, people developed what they called contextual dependencies, which they subdivided into three categories: *intentional stimuli*, which consist of explicitly identified contextual information necessary to successfully perform a task; *incidental stimuli*, which are not explicitly identified as crucial to task performance but have the potential to become associated with particular responses because of their selective presence in the training environment; and *irrelevant stimuli*. In another study, Wright and Shea (1994) examined the intentional and incidental associations a little closer and found that sources of incidental information are contained in the higher order cognitive representation of a task and that these exert a subtle influence on performance. They stress that information sources defined as incidental should not continue to be ignored and deserve future experimental attention. Anderson (1997) re-evaluated some of Wright and Shea's findings and found that there had been a flaw in their experimental set-up. Anderson concluded that indeed there was such a thing as an incidental stimulus, but that its role in motor learning was considerably smaller, although still not negligible. All in all, it seems that learning, whether procedural or declarative, is always context dependent to some extent. Although for contextual cueing to occur an individual's awareness doesn't seem to be essential (e.g., Kimbrough, Wright an Shea, 2001; Roediger, 1990), context effects are indeed larger when the context is noticeable (Davies and Thompson, 1988). Additionally, weakened stimulus-response relations between intentional stimuli and performance cause an increase in the relation between incidental stimuli and performance (Kimbrough et al, 2001). This seems to suggest that encoding contextual cues for motor learning is inherently automatic and *necessary*.

With the apparent context dependency of motor learning, what can be said about skill acquisition through VR? As noted, people tend not to discriminate between what is real and what is made to seem real, so perhaps the principles of learning are identical for both virtual and real. Research with VR training systems has already shown that learners use contextual incidental cues to support and mediate their actions (Lathan, Tracey, Sebrechts, Clawson & Higgins, 2002; Stanney, Mourant & Kennedy, 1998) and that transfer from a VR training simulator to an actual task appears to reduce as tasks become more complex, indicating attentional compensation and a consequent greater reliance on incidental cues (Stanney, 1998).

1.3. The Tower of London

The Tower of London task requires that a number of differently coloured rings or balls placed in a certain start configuration over a certain number of pegs, be moved to an alternate given end configuration of rings in a minimal number of steps and as fast as possible. It was originally devised by Shallice (1982) as a variation on the Towers of Hanoi task, to test the impairment of planning ability in patients with frontal lobe lesions. The reason for creating The Tower of London, with its specific problem space – the graphic representation of the moves possible under the rules of the task –, was "to produce a graded difficulty test". Apparently, Towers of Hanoi, with its restricting rule of only allowing smaller ring placed over larger, was not easily scaleable. Through the years, experimenters have thought of many ways to improve Tower of London or adapt it to fit their specific experimental goals. Some examples are the 'Tower of London Drexel' (e.g. Culbertson, Moberg, Duda, Stern & Weintraub, 2002; Culbertson & Zillmer, 1998) and the Tower of London extended version (Raizner, Song & Levin, 2002). The version of the Tower of London used in this experiment is again very different from that of Shallice and is most like a variation of the Tower of London devised by Ward & Allport (1997), which is more or less a hybrid between Towers of Hanoi and the Tower of London. See figure 1 for some of the variations discussed.

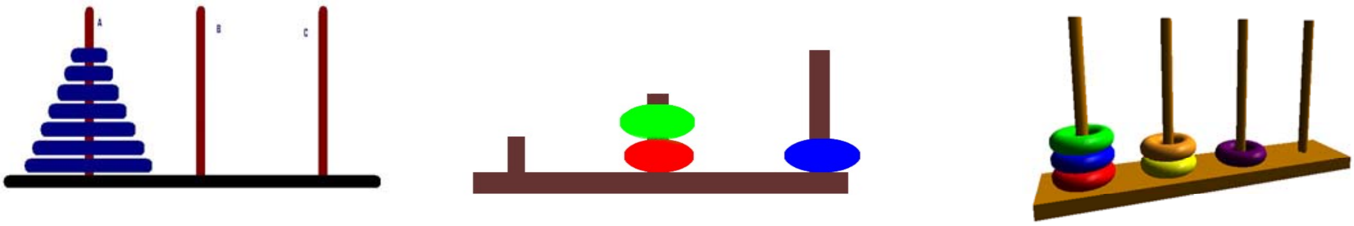

Figure 1: From left to right: Towers of Hanoi, Shallice's Tower of London, and the current version

1.4. The present experiment

The aim was to start off a series of experiments to see *if, to what degree* and exactly *which* contextual factors are in fact relevant to acquiring and transferring motor skill from virtual training environments. The present study served as a pilot study to determine whether such effects even exist by assessing the effect that background colour has on skill acquisition and transfer of training on performance on a virtual Tower of London game. Participants practiced the skill of manipulating virtual rings using a mouse under an altered perspective on the virtual Tower of London game and an unusual correspondence between mouse movement and movement of the manipulated virtual object. This task was made to be relatively difficult, so as to ensure the use of contextual information for encoding motor skill. After some time for training, participants executed a set of twenty-four tasks where cognitive (not motor!) difficulty of the tasks was heightened by an increase in the minimal number of steps needed to complete them. During the last twelve of these cognitively more difficult tasks (i.e. the transfer session), the background changed for the experimental group. The movement times for the rings in the first twelve difficult problems (i.e. the control session) could then be compared with those in the transfer session.

A virtual Tower of London game can be programmed to have users execute some controlled mouse-movement or key presses in order to manipulate the rings and allows for a programmable manipulation of incidental cues such as background colour. There were several reasons for using Tower of London instead of some other task. The primary reason was that the task is two-fold: There is the motor aspect of moving the rings (using a mouse) and the cognitive aspect of solving the presented problem. This has some great advantages: A) the cognitive aspect of the tasks can be increased in difficulty when some incidental cue for motor skill is taken away, so that people cannot compensate for a loss of an incidental cue by allocating more cognitive resources to the motor task, but instead have to focus it on solving the problem; and B) the motor aspects of Tower of London tasks can be made as difficult as one wants them to be. The latter is convenient considering that changing contextual cues appears to affect performance only for more difficult tasks (Anderson, Wright & Immink, 1998; Dibbets, Maes & Vossen, 2002; Speelman & Kirsner, 2001; Wright & Shea, 1991). Other reasons for using the Tower of London were that it was not likely to have been practiced by the participants before, and that it is easy and cheap to obtain a real-life version of the game.

In the present study, participants practiced the skill of manipulating virtual rings using a mouse under an altered perspective on the game and an unusual correspondence between mouse movement and movement of the manipulated virtual object. This skill was made to be relatively difficult, so as to ensure the use of contextual information for encoding, *if* in fact such encoding would ensue. After some time for training, participants executed another twenty-four tasks for which cognitive difficulty of the problems was heightened (the minimal number of steps needed to solve them was increased). At the onset of the last twelve more difficult tasks, the transfer session, the background changed for the experimental group, but remained the same for the control group. The movement times for the rings in the first twelve difficult problems, the control session, could then be compared with those obtained with the last.

The dependent measures the Tower of London software outputs are relatively simple. The program keeps track of the movement time for each ring, the mean and total time to move rings, and the number of moves needed to solve each problem. Because this was just a preliminary study, the investigation was limited to the transfer of perceptual-motor skill, rather than cognitive skill as well. This meant that the only dependent measure relevant to this study was the movement time for each separate ring.

2.Method

2.1. Participants

Twenty-four participants were involved in the experiment. All but a few were acquaintances of the experimenter. Most were students. Participants were aged between eighteen and forty-three with a mean age of 22.4 and an SD of 4.9 years. They had normal or corrected to normal vision. Participants were contacted through telephone or e-mail. They had had no mentionable previous experience with Tower of London tasks.

2.2. Materials

The experiment was executed on an Intel Pentium CPU, 2.80 GHz, 504 MB RAM computer running Windows XP and a Philips17” 10T5/F5/S5 colour monitor. All manipulations in the virtual task environment were executed using a Logitech optical mouse with chord. The computer monitor was situated on a desk at a distance of approximately 50 cm from the participants eyes. Participants were seated on a comfortable rotating desk chair with adjustable height.

The software used was The Tower of London research tool version 1.0., developed at the university of Twente. Because the program did not allow for a full screen interface, the screen was adjusted to a 1024 x 768 pixel resolution to optimize the size of the interface. Because this meant there was still a partly visible desktop around the interface’s edges, the desktop was cleared from icons and made entirely and evenly blue.

2.3.Task

Within the Tower of London interface (see figure 2 on the next page), rings had to be moved using a mouse. Two aspects of the game had however been configured in such a way that difficulty of moving the rings was increased. A first characteristic of the task was an unusual perspective on the game: the virtual game had been rotated 20°, 20° and 40° over its x-, y-, and z-axis respectively. A second characteristic was an altered mouse-angle. This held that the direction in which the manipulated objects (i.e. rings) were moved, deviated at a 45° angle from the direction in which the mouse and cursor were moved (e.g. to move a ring up exactly vertical, the mouse had to be moved at 45° angle left from vertical). The task consisted of two separate parts. First, participants trained the motor-aspect of the task. This involved seventy minutes of non-stop practicing on Tower of London tasks, each solvable in two ring movements. This way, participants trained mostly the motor-aspect of the task, but also gained some insight into the structure of the problem space – though a two step problem is solvable almost immediately. All practice tasks were randomly generated by the Tower of London software. The characteristics of both a peculiar perspective and mouse-angle meant that for all participants, the motor aspect of the task was new, and that any skill developed over the course of the experiment would reflect mainly practice on the Towe of London task, rather than something like previous experience with mouse handling.

After seventy minutes of training, participants were considered to have become sufficiently skilled at manipulating the rings in the awkward game settings. (Though chosen rather arbitrarily, a pilot test showed dramatic improvement over the course of seventy minutes.) In the training phase of the experiment, average movement times were 29228 ms per ring for the first ten rings; 5678 for the last.

Before the practice phase, participant had already been subdivided into two groups; one using a grey background to the tasks, the other a red background. (though both groups had an approximately 4 cm wide blue rim around the edges of the game interface due to the impossibility of a full screen mode.)

The training session was followed by a control session. All participants accomplished twelve harder Tower of London tasks, each solvable in eight steps instead of two. All other aspects of the task and its interface remained the same. For half of the participants from both the blue and red interface conditions the background colour then changed from either red to grey or grey to red at the onset of the transfer session, which consisted of another twelve tasks. These participants constituted the experimental group. For the other participants no background colour change occurred. These constituted the control group. (Make sure not to confuse control *task* with control *group.)* The control and transfer tasks were all configured by the experimenter to be of similar difficulty and identical for all the participants. The reason for control and transfer tasks being more difficult problem solving-wise was that this would leave participants with little cognitive resource left to compensate for any drop in performance caused by a shift in background colour and, thus, a loss of contextual cue. All control and transfer session tasks had two qualities in common to make sure they were as equal in difficulty as possible: They were all solvable in a minimum of eight steps and they were configured to have similar flatness. Flatness refers to whether a start configuration is either shaped like a tower, with most rings being on one peg and the other pegs being empty (which allows for a quicker analysis of the steps needed for solving the problem), or with rings divided over several pegs (a flat configuration), which results in a less clear problem space. In flat configuration, participants have to make more indirect moves (i.e. they more often have to remove a ring before another can be placed in that position).

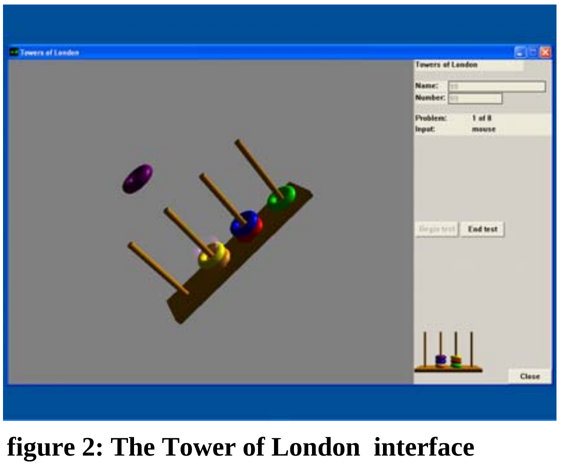

figure 2: The Tower of London interface

2.5. Procedure

All experiments were carried out in an artificially lit cubicle. Upon entering the cubicle, participants were instructed to turn off their mobile phones, be seated and adjust their seat for optimal comfort. Participants were informed that they were to go through seventy minutes of training with simple Tower of London problems and that they were to become proficient at manipulating the rings under the altered perspective and manipulated mouse-angle. They were told that after training on the two step problems, they were to complete another twenty-four tasks of greater difficulty. Participants were instructed to move the rings only when they knew where they were going to place them, so as to avoid movement times partly reflecting solving the problem rather than just motor skill. The practice tasks were randomly generated two-step problems that were divided over a number of directories each containing ten problems. This meant that participants opened a directory, started up the game by clicking the executable, entered their name and number, dragged the game interface to the centre of the screen and pressed a start button before they started with a series of ten problems. After completing ten problems, the game indicated that it had been completed. Participants then clicked a ´close´ button on the interface, went to the next directory - directories were numbered 1 through 25 - and went through the process of generating and executing another ten tasks. The effect that this experimental inefficiency might have had on the outcome will be discussed later. Participants did not use a mouse mat, because their erratic movements – especially in the beginning – covered more area than any mouse mat. Because it was an optical mouse it worked fine nonetheless. Also, participants were not allowed to compensate for the altered mouse-angle by holding the mouse in a way different from what they were used to.

After exactly seventy minutes, participants were told that the practice session had ended. Within this time frame, the slowest person had managed to complete 76 problems, the fastest completed 230. 163 was the average number of completed problems. The average number of rings moved within this time was 359. Participants from the experimental condition were instructed that next, they were going to be completing another twenty-four problems divided over three directories of eight, this time each problem being solvable in a minimum of eight steps, but with all other aspects of the problems remaining the same *up until the thirteenth problem*. They were told that upon the thirteenth problem, the background colour would switch from either grey to red or vice versa (depending on the colour they had been training with), but that they were to disregard this and just keep on solving the problems as fast as they could and within a minimum number of steps, just as they had done the previous problems. Participants were informed of the background colour switch to avoid any effect of startle on movement times. The control group received similar instructions, but no mention of a changing background colour was made since there wasn't going to be any.

3. Results

First, there were no mentionable differences between sessions and colour conditions in the number of steps it took participants to complete each session. Because, nonetheless, there *were* slight differences, some ring movements were discarded to simplify analysis of the data somewhat. The number of ring movements needed for completion of the control session were cut off at the 97^{th} moved ring, for this was the number of rings moved by the participant that had made the fewest erroneous movements to complete the first twelve tasks. The 97^{th} ring movement served as a pivot for analysing the effect of the transition from one background colour to another, for all participants. The number of steps needed for completion of the transfer session- after the colour shift for the experimental group - were cut off at the 101^{st} ring. Of the first 97 ring movements, the first one was discarded, as were the last two of the

last 101 ring movements. This was done so that both sessions could be cut up in three exact (sub-)sessions consisting of equal numbers of rings for more thorough analysis of time (session) effects.
For both sessions, before and after colour shift, 98% one-sided confidence intervals of the movement times were calculated separately. Furthermore, these confidence intervals were calculated for each participant separately, so as not to discard data on account of personal differences. All outliers above the accepted values for the confidence intervals were discarded. For clarification: All results discussed below concern only ring movement time, which is the dependent variable of interest.

For both sessions, mean scores were calculated for each participant. Data were analysed using a 2 (session) x 2 (group) MANOVA. The effect of *session* was significant, $F(1, 22) = 4.68$, $p < .05$ for the separate groups. The effect of session x group, however, was not: $F(1, 22) = 1.34$, $p = .259$. Figure 3 shows a plot with estimated marginal means for both sessions.

Trying to see whether the effect of a background colour shift might be more subtle and of limited duration, the two sessions were cut up into two sessions of three, totalling six sessions, colour shift occurring at onset of the fourth. Data was analysed similarly using a 6 (session) x 2 (group) MANOVA. Again, within-subject effects of sessions were significant: $F(5, 18) = 3.707$, $p = .018$, while between-subject effects of session x group were not.

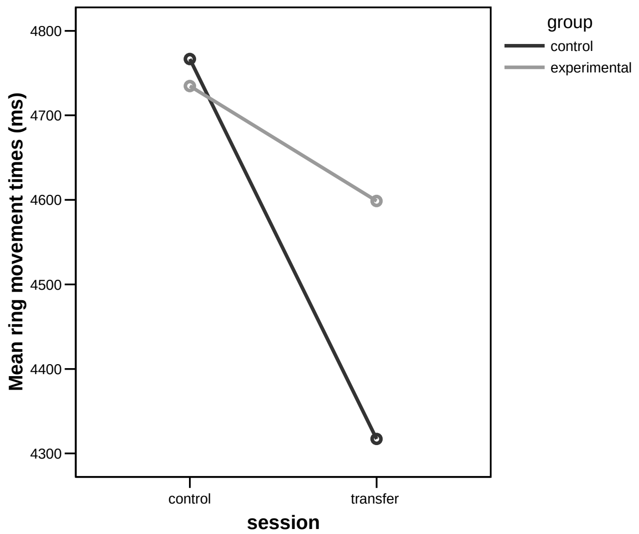

figure 3: Mean ring movement times for control and transfer sessions

Looking at the graph belonging to this data (see figure 4), we do however see what seems like an interaction between sessions three and four. Analysing the two sessions separately in a 2 (session) x 2 (group) MANOVA yields no significant effects however.

No effects of colour change are significant. The next step was to see what can be said about colour sequences. That is, whether grey to red transitions resulted in other effects than red to grey transitions did. Figure 5 yields some insight. The slope for the different colour conditions within the groups are almost identical, although their position differs somewhat. This seems to suggest that there was some structural difference between the experimental and control group. No effects of colour sequence reached statistical significance, however.

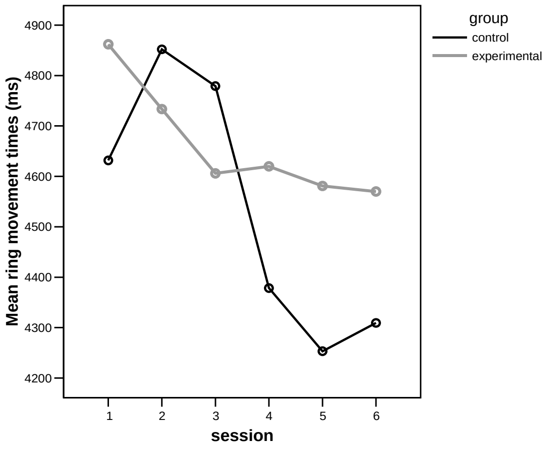

Figure 4: mean movement times for all Participants, calculated for six sub-session Separately

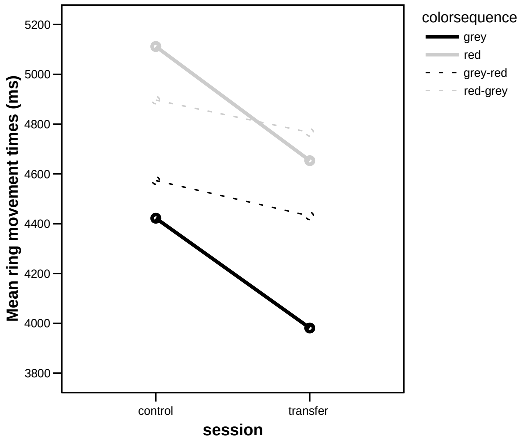

Figure 5: mean ring movement times for control and transfer sessions, calculated for each colour condition separately

A final examination of the data shows that during the control and transfer sessions, people seemed to still be developing skill. Ring movement times have not reached a floor level. A trend line over a scatter plot reflecting averaged movement times for each ring shows a clear learning trend (see figure 6).

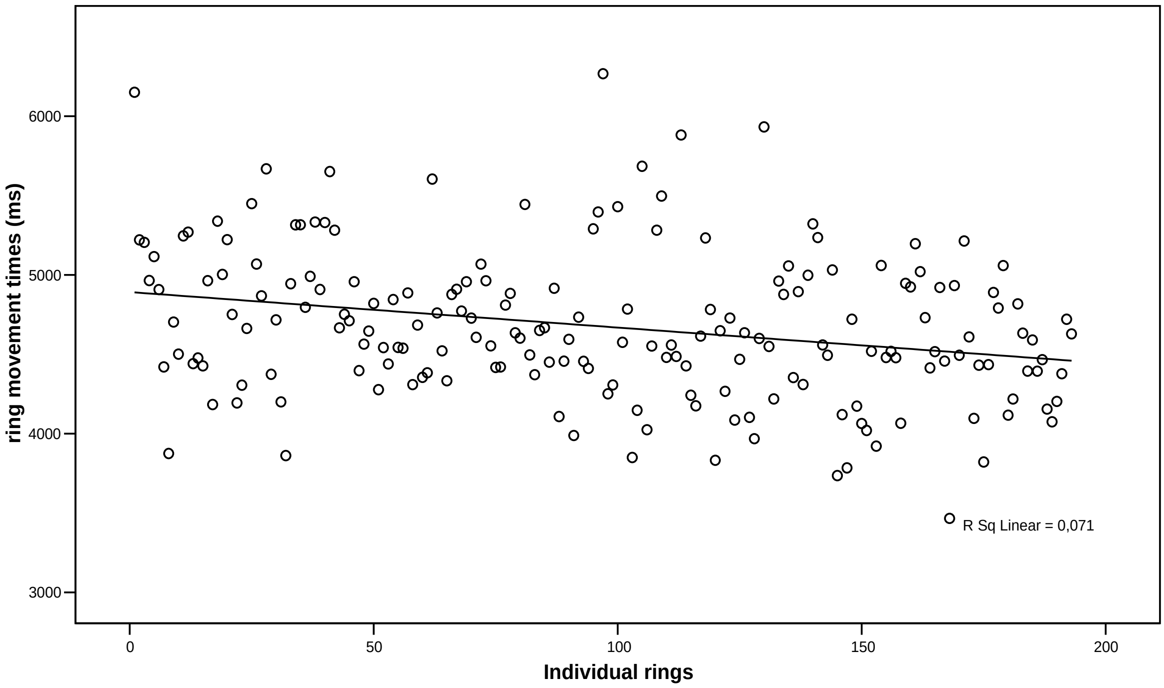

figure 6: A learning trend in movement times during control and transfer sessions

4. Discussion

In trying to find an effect of a contextual dependency on background colour in the development and transfer of skill, no significant effects of colour change between control and experimental groups were found. Only the within-subject variable of session (test time) resulted in significant differences, which probably only reflects that participants were still improving the motor-skill ivolved in moving the rings during the control and transfer sessions. The main hypothesis that “changing the background colour to a virtual Tower of London task after a certain time for traing will negatively affect motor skill, due to developed contextual dependencies” can however not be refuted.

One reason for not finding the intended effects might have to do with the experimental setup. It might, for instance, be possible that the task – moving rings – was just not difficult enough for it to elicit noticeable effects of contextual dependencies. Determining what is hard enough seems tricky. Finding the base line for difficulty/context dependence trade-offs may differ for different task. It seems futile to try and find this baseline for every kind of task imaginable, so to play it safe in testing for these effects, it may be advisable to have participants train *very* difficult tasks. Another possibility is that background colour simply is not an incidental cue, but rather irrelevant (at least for this specific motor task), though considering some of the - albeit non-significant – trends found in the data, this conclusion seems especially unjust. An analysis of the data across sessions shows a slope that doesn´t reach a floor level before the end of the experiment. Participants were clearly still improving their skill during the transfer trials. Although participants had become much better at moving the rings during the training session, they were not done learning after seventy minutes. This does not mean that they could not have developed contextual dependencies, but it could mean that the contextual dependencies simply weren’t strong enough yet for there to be any significant effect.

As stated, the results from the experiment do not justify an abandonment of the main hypothesis. Non-significant trends in the data are cause for hope for finding the contextual cueing effects of background colour after all, especially after considering some of the flaws in the used software. On top of that, there were some initial errors in thinking about why Tower of London was suitable for this experiment, because, as it turns out, it is not.

5. A critical review of the Tower of London

First and foremost: The Tower of London software used for this experiment was far from being suitable for any serious scientific research. Collision detection (i.e. virtual objects behaving like real objects in that they cannot move through and experience resistance from other objects) was flawed. Rings could be moved through the bottom 'wooden' board of the virtual game. When moved too fast, rings would fly back to their starting position, forcing participants to pick it up again and inhibiting them from moving the rings as fast as they perhaps could. The game interface could not be set to cover the entire computer screen, making it unnecessarily small and causing a blue rim (the desktop) to be visible around its edges at all times. Another major flaw was that it was impossible to program the configuration file to generate more than ten problems for each game, thus forcing participants to have to close the game and open a new one after every ten problems, re-entering their name and number for every new game. Learning how to work with a manipulated mouse-angle might very well have been interfered with by the fact that participants had to open directories and click buttons continually under a normal mouse angle. To the very least it will have slowed down the learning process.

As for problems with the interface not covering the entire screen: It seems only logical that background colour used as an incidental cue serves even more so as a single cue when it actually is the *only* background colour. The blue rim around the edges was present across groups and sessions, perhaps thus constantly serving as a constant contextual cue itself. The same might apply to the grey column to the right of the game interface where goal-configuration, buttons and input fields were displayed (see figure 2). Not only might the grey and blue serve as constant contextual cues weakening the measured effects; they might also yield some unknown effect by the contrast between those colours and the background colour chosen for the interface. Apart from that, the grey column might simply be a distracting factor. A consideration for improvement of the software might thus be a full-screen mode of only the relevant parts of the interface from the moment a participant hits the 'begin test' button.

Beside the flawed software, there may be some problems with the Tower of London itself, at least when used as an non-clinical experimental tool. There were significant differences between the performances of different participants. Within the seventy minutes allowed for training the motor aspect of the Tower of London, the slowest person had managed to finish 76 problems (i.e. 152 rings moved plus error) whereas the fastest had finished 220 problems (440 rings moved plus error). Such differences in skill are no doubt normal, but it is perhaps fair to wonder if the Tower of London is not a tool in which such marked differences are only enhanced by the nature of the game and may therefore be less suitable for these experimental means than other tools. For solving a Tower of London problem, one has to scan the image of the starting state, scan the image of the end state, has to discriminate colours as well positions, has to decide which ring to move where and which ring to move first (i.e. has to figure out and perhaps even – along the way - *adapt* a strategy) etc. Thus, not only are people likely to differ on at least some of these aspects, but differences found are also not easily explainable due to the multi-facetted nature of the task.

There is a large degree of subjectivity to the difficulty of problems, because people may employ different strategies for solving them. People do not apply mathematical algorithms when tackling a Tower of London task. Whether one is succesful at selecting a good strategy and exactly which strategy is chosen seems to correlate highly with a persons visuo-spatial working memory capacity (Gilhooly, Wynn, Phillips, Logie & Della Sella). These are all things that one person may be better at than the other, and perhaps cause one person to experience effects of exhaustion – a draining of cognitive load - much sooner than the other. Indeed, many of the participants reported to already feel very tired and/or bored after completing their seventy minute training phase. They were *not* happy! Even though Berg & Byrd (2002) discuss this issue and have gone to some length in analysing the Tower of London problem space in terms of strategy and difficulty, much is still unaddressed. Even if the number of steps *and* the start configuration – tower or flat – remain the same across sessions, this still does not cancel out the effects of individual strategy. If a more absolute understanding of the problem space is desired, perhaps it is wiser to repeat this experiment with the Tower of London Drexel or Shallice's Tower of London software, because their problem spaces are less complex, allowing for a better understanding of the necessary and possible strategies involved in solving problems.

For this experiment it was crucial to keep all problems in the control and transfer sessions at the same level of difficulty for all participants, because different problems might cause different outcomes for individual participants. But even one and the same problem can be easy for one person and difficult for another. There is little literature that tackles the different strategies that people may use for solving Tower of London problems. All the literature that does, applies to forms of Tower of London that differ from this one (e.g. Culbertson et al, 2002; Raizner et al, 2002; Shallice, 1982). These are hardly comparable because the problem space, shortest path algorithms and means-end analyses differ too much. It is important to know more about such strategies on this version of Towers of London if we really want to reduce statistical noise due to configurations of Towers of London that do not take all possible strategies into account. What is more, Kafer and Hunter (1997) concluded that an adapted 4-peg Tower of London task was an unreliable and inaccurate tool for measuring planning and problem solving abilities. Ponsford and Kinsella (1992), using again another variant on the Tower of London, concluded that it did not discriminate well between a control group and an experimental group with deficits in executive functions. Although Baker, Segalowitz and Ferlisi (2001) explain the before mentioned differences mainly by pointing out that there was no single method for scoring Tower of London tasks, the question still remains to which 'tower' category the version used for this experiment belongs. This question is important because then a clearer theoretical understanding of its problem space and the possible used strategies can be achieved. Baker et al also stress that partly because of all the different versions of Tower of London out there, it is extremely difficult to assess its validity. As pointed out, this version is more a hybrid between the Tower of London and the Towers of Hanoi, and very similar to the Tower of London put forth by Ward & Allport (1997). For a discussion on this, see Berg & Byrd (2002). In short, there is still too much uncertainty concerning what it is that the Tower of London measures and how. Also; the Tower of London game is very rich (for lack of a better word) visually and problem solving wise. Perhaps then, Tower of London simply produces too much statistical noise for it to be the best tool for measuring the intended effects.

Another possible shortcoming in the Tower of London as an experimental tool is that participants tend to mix tasks up. It frequently occurred that a participant realised he or she had to figure out where to place a ring while they were already moving it. They may while moving have realised that they were moving the ring in the wrong direction, or they may simply have become overconfident of their own skill of quickly solving problems. All these factors contribute to movement time not being an extremely accurate measure. This article joins that of Berg and Byrd (2002) in urging for more research into this phenomenon and its implications for experimental outcome.

One last critique of Tower of London itself: As mentioned earlier, participants need to be instructed not to move the rings *before* knowing where they are to move them to. This is due to the fact that otherwise, time used for scanning the problem space and planning a move will pollute movement times. However, if participants know where to move a ring when they move it, then the difficulty of the cognitive aspect of the task will possibly no longer interfere with a compensation for losing an incidental cue, i.e. background colour, as much as presumed. There is, in fact, very little direct interference. There ís the interference of working memory being addressed continuously when the harder problems are solved, but to what degree this interferes with motor skill is not clear. It again depends on the strategies people use. If, for instance, people think ´one ring at a time´ then interference will be minimal. If they think several steps ahead then interference will probably be greater. The point remains that a possible explanation for not finding significant results might thus be that participants could compensate (to an unknown degree) for a loss of contextual cueing. What the Tower of London software should allow is not just a logging of ring movement times, but of the time before and in between movements as well, for this constitutes the time it takes participants to think about the problem. These thinking times allow experimenters to 'weigh' the movement times. This seems rather complex so perhaps it is wiser to look for a more suitable tool or experimental set-up for measuring effects of contextual cueing.

A final thought on using Tower of London for further experimenting: The argument that Tower of London is a suitable device because it is easy and cheap to obtain a real life version of it, is not valid, at least not for the non–cognitive aspect. The skills needed for moving rings is very different for real-life versions of the Tower of London. Perhaps using multi-modal interaction with head mounted displays and wired glove technologies will solve this problem.

It may also pay off to examine the Tower Of London Test 3.0. computer version mentioned earlier. This is a new standardized Tower of London variation developed by Pyramid productions.

References

Anderson, T. (1997). Contextual dependencies during perceptual motor skill activities: A re-evaluation of Wright and Shea (1991). *Dissertation Abstracts International: Section B: the Sciences & Engineering, 57*(8-B), 5009.

Anderson, T., Wright, D.L., Immink, M.A. (1998). Contextual dependencies during perceptual-motor skill performance: influence of task difficulty. *Memory, 6* (2), 207-221

Baker, K., Segalowitz, S.J., & Ferlisi, M.C. (2001). The effect of differing scoring methods for the Tower of London task on developmental patterns of performance. *Clinical Neuropsychologist, 15*, 309-313.

Barry, I.S., Mcgaghie, W.C., Petrusa, E.R., Lee G.D. & Scalese, R.J. (2005). Features and uses of high-fidelity medical simulations that lead to effective learning: a BEME systematic review. *Medical Teacher*, 27(1) 10-28.

Berg, W.K., & Byrd, D.L. (2002). The Tower of London spatial problem-solving task: Enhancing clinical and research implementations. *Journal of Clinical and Experimental Neuropsychology, 24,* 586-604.

Culbertson, W.C., Moberg, P.J., Duda, J.E., Stern, M.B. & Weintraub, D. (2004). Assessing the Executive Function Deficits of Patients with Parkinson's Disease Utility of the Tower of London-Drexel. *Assessment magazine, 11* (1), 27-39.

Davies, G. &Thompson, D. (1988). *Memory in context: Context in memory*. New York: Wiley.

Dibbets, P., Maes, J.H.R., & Vossen, J.M.H. (2002). Contextual dependencies in a stimulus equivalence paradigm. *Quarterly Journal of Experimental Psychology. Section B, 55* (2), 97-120.

Doerr, H.K., Quinones, M., Didboye, R.L., & Dunbar, B.S. (2000). Increasing performance through simulator training and empirical evaluation. *Journal of Clinical Monitoring and Computing, 16*(1), 49-66.

Gilhooly, K.J., Wynn, V., Philips, L.H., Logie, R.H. & Della Salla, S. (2002). Visuo-spatial and verbal working memory in the five-disc Tower of London task: An individual differences approach. *Thinking and reasoning, 8* (3), 165-178.

Godden, D.R. & Baddeley, A.D. (1975). Context dependent memory in two natural environments: On land and under water. *British Journal of Psychology, 66*(3), 325 331.

Guthrie, E.R. (1952). *The Psychology of Learning: Revised Edition*. Massachusetts: Harper Bros.

Guthrie, E.R. & Powers, F. (1950) .*Educational Psychology*. New York: Ronald Press Co.

Hays, R.T., Jacobs, J.W., Prince, C. & Salas, E. (1992). Flight simulator training effectiveness: a meta analysis. *Military psychology, 4*(2), 63-74.

Henry, F. (1958). Specificity vs. generality in learning motor skills. College Physical Education Proceedings, 126-128.

Hintzman, D.L., (1990). Human learning and memory: connections and dissociations. *Annual review of psychology, 41*: 109-39

Hull, C.L. (1943) *Principles of Behavior*. New York: Appleton-Century-Crofts.

James, L.K., Lin, C., Steed, A., Swapp, D., & Slater, M. (2003). Social Anxiety in Virtual

Environments: Results of a Pilot Study. *Cyberpsychology & Behavior, 6*(3), 237-244.
Jang, D.P., Kim, I.Y., Nam, S.W., Wiederhold, B.K., Wiederhold, M.D. & Kim, S.I. (2002). Analysis of Physiological Response to Two Virtual Environments: Driving and Flying Simulation. *Cyberpsychology & Behavior*, *5*(1), 11-18.
Kafer, K.L., & Hunter, M. (1997). On testing the face validity of planning/problem-solving tasks in a normal population. *Journal of the International Neuropsychological Society*, *3*, 108- 119.
Kimbrough, S. K., Wright, D. L., & Shea, C. H. (2001). Reducing the saliency of intentional stimuli results in greater contextual-dependent performance. *Memory, 9*, 133-143.
Lathan, C.E., Tracey, M.R., Sebrechts, M.M., Clawson, D.H. & Higgins, G.A. (2002). Using virtual environments as training simulators: measuring transfer. In Stanney, K.M. (Ed.), *Handbook of virtual environments: design, implementation, and application* (pp 403–414). New Jersey: Lawrence Erlbaum, Mahwah.
Ma, H.I., Trombly, C.A., Robinson-Podolski, C. (1999). The effect of context on skill acquisition and transfer. *American Journal of occupational therapy*. *53*(2): 138-144.
Mani, S., Shoor, S.K. & Pedersen, H.S. (2004). Experience with a simulator for training ammonia plant operators. *Plant/Operations Progress, 9*(1), 6-10. Retrieved September 1, 2005, from http://www3.interscience.wiley.com/cgi-bin/abstract/109076959/abstract.
Pavlov. I. P. (1928). Lectures on conditioned reflexes: Twenty-five years of objective study of the higher nervous system activity (behavior) of animals. Vol. 1. W. H. Gantt (Trans. & Ed.), New York: International Publishers.
Peugnet, F., Dubois, P. & Rouland, J.F. (1998). Virtual reality versus conventional training in retinal photocoagulation: A first clinical assessment. *Computer Aided Surgery, 3*(1), 20–26.
Ponsford, J.L. & Kinsella G. (1992). Attentional deficits following closed head injury. *Journal of clinical and experimental psychology, 14*, 822–838.
Pyramid Productions. (2005). *Tower Of London Test handleiding 3.0.* Retrieved September 3, 2005, from http://home.hetnet.nl/~fckovacs/newtests.html#TLT
Raizner, R.D., Song J., Levin, H.S. (2002). Raising the ceiling: the Tower of London – extended version. *Developmental Neuropsychology, 21*, 1-14.
Reeves, B. & Nass, C. (1996). *The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places*. New York: Cambridge University Press.
Roediger, H. L. (1990). Implicit memory: Retention without remembering. *American Psychologist, 45*(9), 1043-1056.
Rowe, J. B., Owen, A. M., Johnsrude, I. S., & Passingham, R. E. (2001). Imaging the mental components of a planning task. *Neuropsychologia*, *39*, 315–327.
Shallice, T. (1982). Specific impairments of planning. *Philosophical Transactions of the Royal Society of London, B298*, 199-209.
Speelman, C. & Kirsner, K. (2001). Predicting transfer from training performance. *Acta Psychologica*, *108*, 247-281.
Stanney, K.M., Mourant, R., & Kennedy, R.S. (1998). Human factors issues in virtual environments: A review of the literature. *Presence: Teleoperators and Virtual Environments*, *7*(4), 327-351.
Tulving, E. & Osler, S. (1968). Effectiveness of retrieved cues in memory for words. *Journal of experimental psychology, 77*, 593-606.
Tulving, E. & Thomson, D.M. (1973). Encoding specificity and retrieval process in episodic memory. *Psychological review, 80*, 352-73.
Ward, G., & Allport, A. (1997). Planning and problem solving using the 5-disc Tower of London task. *Quarterly Journal of Experimental Psychology, 50*, 49-78.

White, C.R., Carson J.L., Wilbourn, J.M. (1991). Training effectiveness of an M-16 rifle simulator. *Military Psychology, 3*(3), 177-184.

Wright, D. L. & Shea, C. H. (1991). Contextual dependencies in motor skills. *Memory and Cognition, 19*(4), 361-370.

Wright, D. L. & Shea, C. H. (1994). Cognition and motor skill acquisition: Contextual dependencies. In C. R. Reynolds (Ed.), Cognitive assessment: A multidisciplinary perspective (pp 89-106). New York: Plenum Press.