Can Self-Explanations and Thinking Aloud Promote Performance at the Tower of Hanoi Problem?

Christine Blech

FernUniversität in Hagen

Abstract. An experiment (*N* = 89) with a computerized version of the Tower of Hanoi (ToH) investigated the combined effects of instructed verbalization and self-reported expertise on problem solving performance in a 2 (ToH experts vs. novices) × 3 (verbalization instruction) between-subjects design. The hypothesis that self-explanations can promote ToH performance as compared to silent thinking or conventional thinking aloud was confirmed for ToH experts, but not for novices. Experts who had to explain and justify each move needed more time, but fewer moves to complete the ToH problems than silent or conventional aloud-thinkers. In contrast, the novices needed more time and more moves when engaging in self-explanations. Based on the Cognitive Load Theory, the conflicting demands of verbalization, schema acquisition, and goal-oriented solution behavior are discussed, explaining the novices' inefficient solutions and their presumed working memory overload. Concerning the experts' performance, the positive impact of mental anticipation is highlighted.

One hundred years before cognitive psychologists in Germany discussed the controversial question of whether and how intelligence affects complex

Blech, C. (2025). Can Self-Explanations and Thinking Aloud Promote Performance at the Tower of Hanoi Problem?. In A. N. Wendt, D. V. Holt & L. von Stockhausen (Eds.), *Komplexität und Problemlösen. Festschrift für Joachim Funke zum 70. Geburtstag* (S. 35–46). Heidelberg University Publishing. https://doi.org/10.17885/heiup.1067.c23280

problem solving (Dörner et al., 1983; Funke, 1983), in 1883, the French mathematician Édouard Lucas (see Hinz et al., 2018) worked out a puzzle that would tease the brain of both test subjects and researchers even two turns of centuries later. The Tower of Hanoi (ToH) was born. To psychologists, the ToH is known as a well-defined, simple, versatile problem solving paradigm with applications in research fields as diverse as child psychology (Resing et al., 2020), psychology of old age (Zinke et al., 2014), clinical psychology (Knapp et al., 2017), neurosciences (Anderson et al., 2005), and, of course, general and experimental cognitive psychology (e.g., Welsh & Huizinga, 2005).

The Tower of Hanoi: Problem or Routine Task?

The ToH does not require elaborate equipment; the rules are plain and clear (see, e.g., Simon, 1975): Given three pegs and a varying number of disks, different in size, a pyramid-like stack of disks has to be transferred from one peg to another. Only one disk, a disk from the very top of a stack, may be moved at a time. A larger disk may never be placed on top of a smaller disk. Besides this original "tower-to-tower" version of the ToH, there are also variants in which the initial state and/or the goal states are in "flat" positions or display partial towers (for examples see, e.g., Anderson et al., 2005, or the Method section below). While the haptic, physical variant of the ToH is suitable for studies with specific samples, e.g., children (Resing et al., 2020), the conceptually equivalent computer-based implementation has some advantages for experiments with healthy adult participants. Sequences of different, single ToH problems can be presented quickly in a standardized manner, correct and wrong moves are recorded automatically along with the solution times (Mataix-Cols & Bartrés-Faz, 2002), and even online testing would be feasible.

Compared to real-life problems, the ToH seems artificial and low in complexity. Following the taxonomy by Dörner (1976), the ToH is a prototypical interpolation problem: the *clarity of the goal state* is high (i.e., no goals need to be defined or negotiated by the problem solver) and the *awareness about available means* to reach the goal is high as well. With sufficient exercise, the problem-defining barrier between given state and goal state (Frensch & Funke, 1995) disappears. That is why in research literature, the tower of ToH is sometimes classified as a measure of planning (e.g., Knapp et al., 2017) or executive functions (Zinke et al., 2014) rather than as a problem solving paradigm. Anderson et al. (2005) trained participants with various ToH subtasks (each with five disks and an optimal move number of 28) to learn the so-called "sophisticated perceptual strategy" in the sense of Simon (1975). After two days, all participants could solve the tasks virtually flawlessly in a predictable manner. This enabled the authors to map distinct, involved brain areas for phases of planning and representation (e.g., encoding positions of disks, setting subgoals), phases of memory activity (e.g., remembering a plan), and phases of motor execution (e.g., moving disks).

On the other hand, the problem solving perspective on the ToH has shown that, as long as participants approach the ToH "naturally", i.e., untrained, differential solution strategies (for a classification, see Simon, 1975) can be identified in behavioral data and self-reports, leading to a considerable amount of individual variation in performance (Welsh & Huizinga, 2005). Sometimes, when solutions reach a dead end and moves have to be reversed, solvers may even experience conflicts (Jostmann & Gieselmann, 2014), similar to the goal conflicts in complex problems (Dörner et al., 1983). So, it is important to consider problem solvers' previous experience with the ToH when assessing their performance.

Thinking Aloud and Self-Explanations

Anderson et al. (2005) mapped problem solving processes in the ToH based on solution times and fMRI analysis. Another approach to access cognitive processes is the method of thinking aloud (van Someren et al., 1994). Thinking aloud means to freely express everything that comes to one's mind while working on a task or—in this case—a ToH problem. The question of whether verbalization affects thinking routines has been a debate for a long time (see Fox et al., 2011, and the reply by Schooler, 2011). According to Fox et al. (2011), the potential reactivity depends on the abstraction and elaboration of the verbalization. In their meta-analysis of 94 studies, so-called concurrent verbalization, conceptualized as vocalizing thoughts which would have been inner speech otherwise, showed no effects of reactivity. In contrast, verbalizing specific descriptions or explanations went along with higher performance than free, silent thinking. In an experiment with the water jar problem by Luchins (1942), Blech et al. (2020) found that concurrent thinking aloud neither changed the quality of solution nor the solution time, although one might guess that expressing thoughts would take more time than thinking in silence.

In the study at hand, the particular focus was on the potentially beneficial effects of the more elaborate verbalization techniques like self-explanation methods (Chi et al., 1989; Neuman & Schwarz, 1998). Problem solving took place under three different verbalization conditions: in silence, with concurrent thinking aloud, and with self-explanations, meaning that while working on the ToH, problem solvers had to explain where they were moving the disks and why. The central hypothesis was that self-explanation would enhance the ToH performance as compared to silent problem solving or problem solving with mere concurrent thinking aloud.[1] Additionally, solution times and potential effects of expertise were investigated, i.e., whether problem solvers who were already familiar with the ToH (a) solved the ToH with fewer moves than novices and (b) might benefit either less or more from self-explanation instructions.

Method

Recruitment and Sample

The experiment was planned and conducted in the context of distance teaching. Fifteen students of the FernUniversität in Hagen recruited six participants each from their personal acquaintances and tested them at the students' homes. Rooms and computers were prepared according to standardized guidelines (see Blech, 2022), e.g., distractors like radios and telephones were muted, computers were in

[1] The data collection was part of a student project in 2014. The hypothesis was set up prior to raising the data, but there was no official pre-registration.

full-screen mode, only one participant was tested at a time. By drawing lots (from a pool of six lots, containing codes for three different experimental conditions, each code twice), every student investigator made sure to assign treatment conditions randomly while still aiming for balanced cell sizes.

Altogether, 91 participants were tested, one of them to replace the data of an excluded participant whose experimental condition was not documented properly. During data analysis, another data set was excluded for being an extreme outlier with respect to solution time (average time per task: more than 5 *SD* above the group mean; average time per move: more than 7 *SD* above the group mean). The protocol suggested major interruptions during the test session. Out of the remaining 89 participants, 54% were female, 46% male. The age ranged from 16 to 65 years, the average being 36.62 (*SD* = 13.29) years. Asked about their familiarity with the ToH on a scale from 1 (not familiar at all) to 5 (very familiar), 63% rated themselves as not very familiar (scorings 1 or 2) and will be referred to as novices in the analyses. The remaining 37% (scorings 3, 4, or 5) rated themselves as at least medium-familiar and will be referred to as experts (see Table 1 for further details).

Table 1. *Distribution of expertise across experimental conditions*

Experimental Condition	Novices		Experts		
	n	%	*n*	%	Total *n*
Silent control	17	57%	13	43%	30
Thinking aloud	23	74%	8	26%	31
Self-explanation	16	57%	12	43%	28
Total	56	63%	33	37%	89

Note. Novices rated themselves at 1 or 2, and experts rated themselves at 3, 4, or 5 on the ToH familiarity Likert scale. Although the frequencies were not perfectly balanced, the test of stochastic dependency between expertise and experimental condition was not significant, $\chi^2(2) = 2.59$, $p = .27$.

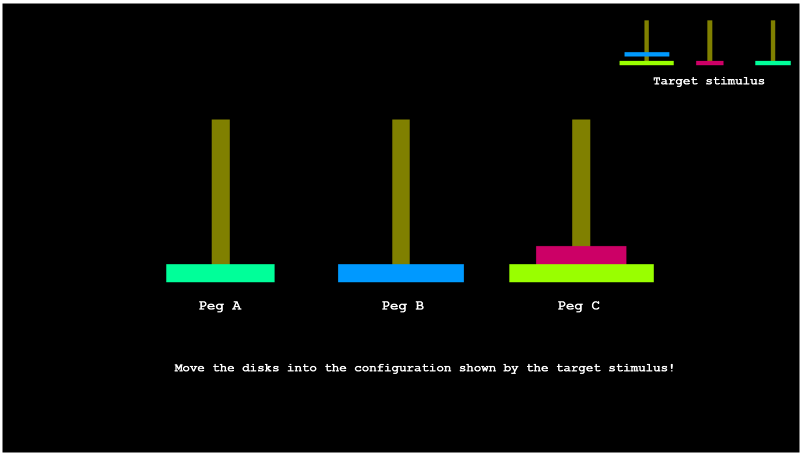

Figure 1. *User interface of the ToH software. Translated to English from the original German version. The disks were moved using a computer mouse. The optimal move number is 8.*

Material

The experimental software (Hertel, 2014), based on C ++, contained ten computerized ToH problems: Two practice problems with four disks and optimal (i.e., minimal) move numbers of 7 and 8, followed by the eight regular problems with four disks and optimal moves from 8 (problem 1) to 15 (problem 8). The problems were "flat-to-flat" towers presented in order of ascending difficulty, i.e., with an increasing number of optimal moves. An example is given in Figure 1; the full software is available at the Open Science Framework project (Blech, 2022). For all problems, the background color was black, text color was white, and the disks of different sizes varied in color to make discrimination between them easier. Disks were moved by using a computer mouse. The ToH standard instruction ran: "Try to reach the target state in as few moves as possible." The rules outlined in the introduction were explained prior to the first practice task: (a) move only one disk at a time, and (b) it is not possible to place a larger disk on top of a smaller disk. A sample sketch of a ToH configuration illustrated the position of actual and target position and how to move disks from one peg to another. After the practice task participants reassured that they had understood the rules. Otherwise, they could ask the experimenter for help before continuing with the main part of the experiment.

Design

Depending on the experimental condition, the ToH standard instruction was complemented by a verbalization instruction manipulated between subjects in a one-factorial design. Participants of the thinking-aloud condition were instructed: "Try to reach the target state in as few moves as possible. Speak out any thoughts that are going through your mind during this time." For participants of the self-explanation condition, the instruction was: "Try to reach the target state in as few moves as possible. Before each move, explain where you are moving a disk and why." Participants of the silent control condition were only given the standard instruction: "Try to reach the target state in as few moves as possible." Conditions were assigned randomly.

Procedure

In the private rooms of the student experimenters, the desks and computers were prepared according to the guidelines mentioned above. The experimenters welcomed the participants by reading off a standard text stating that the study was part of a students' research project in psychology, that the aim was to investigate cognitive processes during problem solving, and that the duration would be approximately 30 minutes. Participants from the think-aloud and self-explanation condition were informed that they had to verbalize, though the verbalizations themselves were not recorded. After signing the informed consent sheet on paper, participants continued with the ToH problems on a computer. All instructions, including the group-specific verbalization instructions, were embedded in the software. Depending on the randomly (per lot) assigned condition, the experimenter started a different version of the software. A glass of water was provided for all participants with and without verbalization instructions. The experimenters stayed in the room with the participant, listening to the verbalizations and making sure that the instructions were followed. If think-aloud or self-explanation participants turned silent for five seconds or more, the experimenter reminded them to speak out their thoughts on solving the problem. After completing the ToH problems, participants had to answer the questions on demographical variables and the rating scale on ToH familiarity. Finally, they were fully debriefed. They read and were invited to keep a handout explaining, among others things, the three experimental conditions and the hypothesis, together with the experimenters' contact details and the opportunity to ask any questions. The experimenters made sure the data was properly saved on their local computers and transferred the files to the supervising teacher of the project.

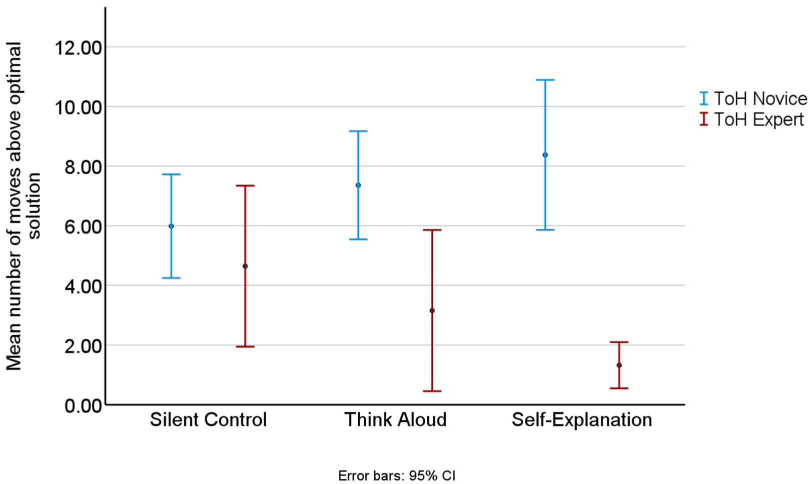

Figure 2. *ToH performance as measured by the average number of moves above the optimum, displayed for silent thinking, thinking aloud, and self-explantation, separated by expertise. Low numbers indicate high performance.*

Measures and Variables

For each ToH problem, the software recorded the number of legal moves and the solution time in seconds exact to three decimal places as raw data. Attempted illegal moves (mistakes) were also recorded though their execution was blocked: If a participant tried an illegal move, the disk was moved back to its previous position. Practice problems were not included in the analysis. The solution time in seconds was averaged over the eight ToH problems. The average number of legal moves above the optimum served as a measure of performance so that ToH problems requiring more moves had the same weight as simpler problems. According to experimenters' protocols, attempted illegal moves sometimes turned out to be affected by problems with the computer mouse (i.e., not clicking properly, releasing the mouse key too early) and were thus dismissed. SPSS version 28.0.1.1 was used for the statistical analyses. Data, along with an analysis script, is uploaded at the Open Science Framework project (Blech, 2022).

Results

ToH Performance

Figure 2 displays the average numbers of moves above the optimum for the three experimental conditions, separated by expertise. A 3 (verbalization) × 2 (expertise) between-subjects ANOVA showed a significant main effect of expertise on performance, $F(1, 83) = 23.74$, $p < .001$, $\eta_p^2 = .22$, the experts needing fewer moves than the novices. While the main effect of verbalization was not significant, $F < 1$, ns., the interaction of verbalization × expertise was significant, $F(2, 83) = 3.90$, $p = .024$, $\eta_p^2 = .09$. Simple effect analyses for the novices did not show a significant main effect for the verbalization condition, $F(2, 83) = 1.61$, $p = .207$, $\eta_p^2 = .04$. In fact, the descriptive results hinted in the direction opposing the hypothesis (see Fig. 2). For the experts, the result pattern matched the expected direction, the number of moves being lowest under the self-explanation instruction and highest for the silent control condition. Yet, the simple effect of verbalization was not significant for the ToH experts either, $F(2, 83) = 2.32$, $p = .105$, $\eta_p^2 = .05$.

Solution times

Paralleling the analyses on the performance measure, a 3 (verbalization) × 2 (expertise) between-subjects ANOVA was conducted, with solution time as the dependent variable. The average solution time for novices was significantly higher than the solution time for experts, $F(1, 83) = 27.80, p < .001, \eta_p^2 = .25$. The main effect for verbalization became significant as well, $F(2, 83) = 8.09$, $p < .001$, $\eta_p^2 = .16$: The longest solution times occurred under the self-explanation condition (see Fig. 3). The interaction of verbalization × expertise was not significant, $F < 1$, ns. Simple effect analyses for the novices showed a significant main effect for the verbalization condition, $F(2, 83) = 7.67$, $p < .001$, $\eta_p^2 = .16$. Unlike with the performance measure, the result pattern for the expert subsample was not opposed to, but, rather, mirrored the result pattern for the novices. Experts, too, had longer average solution times under the self-explanation condition than under the silent control and under the think-aloud condition, although the simple main effect did not become significant, $F(2, 83) = 2.11$, $p = .127$, $\eta_p^2 = .05$.

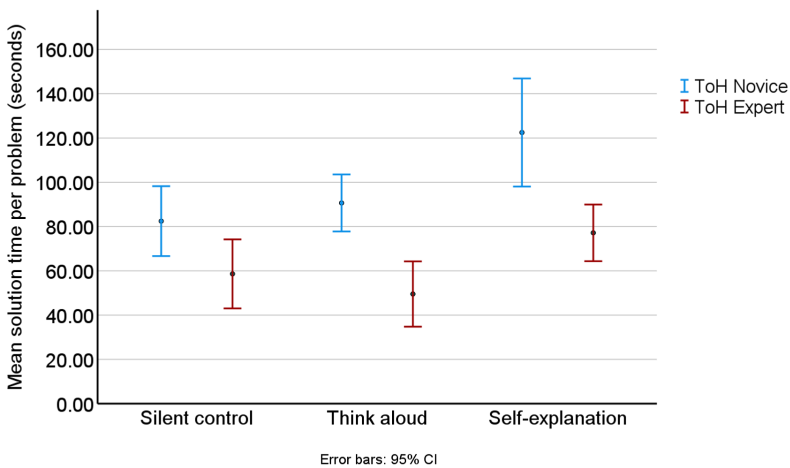

Figure 3. *ToH solution times, displayed for silent thinking, thinking aloud, and self-explanation, separated by expertise.*

Exploratory Analysis

For explorative purposes, solution times were correlated with the performance measure. On the level of single ToH problems, problems requiring more moves were problems requiring more time, with Pearson correlation coefficients from $r(89) = .60$ for Problem no. 7 and $r(89) = .84$ for Problem no. 4. Similarly, aggregated over all problems per participant, the generally quick problem solvers tended to be the generally accurate problem solvers, the correlation between average number of moves and average solution time being $r(89) = .59, p < .001$.

Discussion

The study reported in this chapter investigated the potential impact of self-explanation and thinking-aloud instructions on the solution performance of a computerized version of the ToH, a simple and frequently applied problem solving paradigm. According to their self-reports of familiarity with the ToH, participants were either classified as experts or novices. A clear, yet not surprising, finding was that on average, experts solved the ToH problems with fewer moves and faster than novices. In the total sample, self-explanations were neither advantageous nor obstructive to the solution performance.

Interestingly, prior experience with the ToH moderated the effect of verbalization instructions: Experts seemed to benefit from self-explanations. They took fewer moves under this condition as compared to silent thinking or concurrent thinking aloud without explanations. Meanwhile, their solution times were not reduced, suggesting that the self-explanation instruction brought the experts to thoroughly anticipate and justify each move before executing it. Similar effects have been reported by Fansher et al. (2022), who compared problem solving of the ToH in a traditional computerized version with a mental imagery condition. The latter condition forced participants to internally simulate the problem solving sequence instead of simply executing the solution step by step, including trial-and-error procedures. As the authors note, in line with the Cognitive Load Theory (Sweller, 1988), mental simulation may consume working memory capacity that would otherwise have been available for different cognitive processes, such as acquiring new problem solving strategies or schemas. The same effect might occur when forced self-explanations occupy working memory.

Working memory capacity, in turn, is one crucial factor distinguishing experts from novices: By encoding not more, but larger chunks of information (Chase & Simon, 1973) and by using more efficiently organized higher-order problem knowledge (Chi et al., 1982), experts can still have free capacities in problem solving situations that make novices struggle. In the reported experiment, ToH novices performed worst under the self-explanation condition and best under the silent thinking condition. Self-explanations prolonged the solution process but did not improve its outcome. As an interpretation, it is suggested that novices' working memories were overloaded with the combined demand of generating problem solving strategies and putting effort into verbalized self-explanations. As Chi et al. (1989) report, even in traditional thinking-aloud settings, not being prompted to explain solution steps, good problem solvers tend to provide abstracting self-explanations spontaneously, whereas bad problem solvers stick to the mere descriptive level more often. Sweller (1988) puts forward the idea that goal-directed problem solving and the acquisition of knowledge in terms of solution schemes are conflicting processes. Although, once acquired, knowledge about problems is helpful in order to achieve goal states (e.g., Funke, 1992, on complex problem solving), at earlier stages of solving an unfamiliar problem, knowledge acquisition through (free) exploration can be considered a costly investment. Likely, the novices suffered from this dilemma to allocate their cognitive resources appropriately.

Since the verbalizations were not recorded, it remains a little speculative whether the experts' self-explanations were indeed different in quality from the presumably descriptive and less efficient self-explanations of the novices. This methodological limitation could be addressed in a replication, which could also aim for a more balanced design with equally large groups of experts and novices, randomly assigned to each verbalization condition. It would be interesting to investigate linguistic and syntactic features of the verbalizations, possibly detecting more elaborate, well-structured phrases under self-explanation conditions, indicating effortful cognitive processing. Another future perspective would be an in-depth analysis of the solutions paths (see, e.g., Hinz, 2012), which was beyond the scope of this work.

One fundamental criticism of the ToH concerns the paradigm's artificial nature and its low complexity (cf. Dörner, 1981). Betsch et al. (2011, p. 1993) compare the ToH in cognitive psychology to the fruit fly (*Drosophila melanogaster*) in biological laboratory research. The question is: What can we learn about humans from flies? What can we learn from simple, controlled laboratory experiments with the ToH about complex real-life behavior? Against the background of the present results, one basic lesson would be that expertise matters. Gaining experience takes time and effort, but once acquired, it allows experts to have a relatively clear view of the aspects relevant to the solution of a problem. Of course, like novices, experts should be given time for reflection instead of being rushed into a critical decision. There is no doubt that problem prevention and forward planning are more advisable than frantic reactions in a crisis which has gotten out of control. The warning of Dörner et al. (1983) regarding "intellectual emergency reactions" seems still very much up to date, forty years later. Self-explanations or more creative, elaborate variations of this technique can guide experts' problem solving processes, helping to avoid purely intuitive or trial-and-error reactions and fostering mental simulations of several different—both obvious and counterintuitive—solution approaches.

Acknowledgements. I would like to thank the students from summer term 2014 for recruiting and testing participants as well as for their constructive comments on planning and pretesting the experiment.

References

Anderson, J. R., Albert, M. V., & Fincham, J. M. (2005). Tracing problem solving in real time: fMRI analysis of the subject-paced Tower of Hanoi. *Journal of Cognitive Neuroscience, 17*(8), 1261–1274. https://doi.org/10.1162/0898929055002427

Betsch, T., Funke, J., & Plessner, H. (2011). *Allgemeine Psychologie für Bachelor: Denken – Urteilen, Entscheiden, Problemlösen* [General psychology for bachelor: Thinking judgment, decision making, problem solving]. Springer. https://doi.org/10.1007/978-3-642-12474-7

Blech, C. (2022). *Thinking aloud, self-explanation, and the Tower of Hanoi* [OSF Project]. https://osf.io/kz35c/

Blech, C., Gaschler, R., & Bilalić, M. (2020). Why do people fail to see simple solutions? Using think-aloud protocols to uncover the mechanism behind the Einstellung (mental set) effect. *Thinking & Reasoning, 26*(4), 552–580. https://doi.org/10.1080/13546783.2019.1685001

Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology, 4*(1), 55–81. https://doi.org/10.1016/0010-0285(73)90004-2

Chi, M. T. H., Bassok, M., Lewis, M., Reimann, P., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science, 13*, 145–182. https://doi.org/10.1207/s15516709cog1302_1

Chi, M., Glaser, R., & Rees, E. (1982). Expertise in problem solving. In R. Sternberg (Ed.), *Advances in the psychology of human intelligence* (pp. 7–79). Erlbaum.

Dörner, D. (1976). *Problemlösen als Informationsverarbeitung* [Problem solving as information processing]. Kohlhammer.

Dörner, D. (1981). Über die Schwierigkeiten menschlichen Umgangs mit Komplexität [The difficulties of human encounters with complexity]. *Psychologische Rundschau, 32*(3), 163–179.

Dörner, D., Kreuzig, H. W., Reither, F., & Stäudel, T. (1983). *Lohhausen: Vom Umgang mit Unbestimmtheit und Komplexität* [Lohhausen: Problem solving in uncertain and complex problems]. Hans Huber.

Fansher, M., Shah, P., & Hélie, S. (2022). The effect of mode of presentation on Tower of Hanoi problem solving. *Cognition, 224*, 105041. https://doi.org/10.1016/j.cognition.2022.105041

Fox, M. C., Ericsson, K. A., & Best, R. (2011). Do procedures for verbal reporting of thinking have to be reactive? A meta-analysis and recommendations for best reporting methods. *Psychological Bulletin, 137*(2), 316–344. https://doi.org/10.1037/a0021663

Frensch, P. A., & Funke, J. (Eds.). (1995). *Complex problem solving: The European perspective*. Lawrence Erlbaum Associates, Inc.

Funke, J. (1983). *Einige Bemerkungen zu Problemen der Problemlöseforschung oder: Ist Testintelligenz doch ein Prädiktor* [Some comments on the problems in problem-solving research, or: Is test intelligence still a predictor]? *Diagnostica, 29*, 283–302.

Funke, J. (1992). *Wissen über dynamische Systeme: Erwerb, Repräsentation und Anwendung* [Knowledge about systems: Acquistion, representation, and application]. Springer. https://doi.org/10.1007/978-3-642-77346-4

Hertel, J. (2014). *FlexibleDesign* [Unpublished computer software].

Hinz, A. M. (2012). Graph theory of tower tasks. *Behavioural Neurology, 25*(1), 13–22. https://doi.org/10.3233/BEN-2012-0345

Hinz, A. M., Klavžar, S., Milutinović, U., & Petr, C. (2018). *The Tower of Hanoi. Myth and math* (2nd ed.). Birkhäuser.

Jostmann, N. B., & Gieselmann, A. (2014). When you have to climb downhill to reach the top: The effect of action versus state orientation on solving a goal-subgoal conflict in the Tower of Hanoi task. *Experimental Psychology, 61*(5), 394–401. https://doi.org/10.1027/1618-3169/a000260

Knapp, F., Viechtbauer, W., Leonhart, R., Nitschke, K., & Kaller, C. P. (2017). Planning performance in schizophrenia patients: A meta-analysis of the influence of task difficulty and clinical and sociodemographic variables. *Psychological Medicine, 47*(11), 2002–2016. https://doi.org/10.1017/S0033291717000459

Luchins, A. S. (1942). Mechanization in problem solving: The effect of Einstellung. *Psychological Monographs, 54*(6), i–95. https://doi.org/10.1037/h0093502

Mataix-Cols, D., & Bartrés-Faz, D. (2002). Is the use of the wooden and computerized versions of the Tower of Hanoi puzzle equivalent? *Applied Neuropsychology, 9*(2), 117–120. https://doi.org/10.1207/S15324826AN0902_8

Neuman, Y., & Schwarz, B. (1998). Is self-explanation while solving problems helpful? The case of analogical problem-solving. *British Journal of Educational Psychology, 68*(1), 15–24. https://doi.org/10.1111/j.2044-8279.1998.tb01271.x

Resing, W. C. M., Vogelaar, B., & Elliott, J. G. (2020). Children's solving of 'Tower of Hanoi' tasks: Dynamic testing with the help of a robot. *Educational Psychology, 40*(9), 1136–1163. https://doi.org/10.1080/01443410.2019.1684450

Schooler, J. W. (2011). Introspecting in the spirit of William James: Comment on Fox, Ericsson, and Best (2011). *Psychological Bulletin, 137*(2), 345–350. https://doi.org/10.1037/a0022390

Simon, H. A. (1975). The functional equivalence of problem solving skills. *Cognitive Psychology, 7*(2), 268–288. https://doi.org/10.1016/0010-0285(75)90012-2

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12*(2), 257–285. https://doi.org/10.1207/s15516709cog1202_4

van Someren, M. W., Barnard, Y. F., & Sandberg, J. A. C. (1994). *The think aloud method: A practical guide to modelling cognitive processes*. Academic Press.

Welsh, M. C., & Huizinga, M. (2005). Tower of Hanoi disk-transfer task: Influences of strategy knowledge and learning on performance. *Learning and Individual Differences, 15*, 283–298. https://dx.doi.org/10.1016/j.lindif.2005.05.002

Zinke, K., Zeintl, M., Rose, N. S., Putzmann, J., Pydde, A., & Kliegel, M. (2014). Working memory training and transfer in older adults: Effects of age, baseline performance, and training gains. *Developmental Psychology, 50*(1), 304–315. https://doi.org/10.1037/a0032982