When Do Nonspecific Goals Help Learning? An Issue of Model Quality

Saskia Kistner (kistner@paed.psych.uni-frankfurt.de)
Institute of Psychology, Goethe University, Theodor-W.-Adorno-Platz 6, D-60323 Frankfurt/Main, Germany

Bruce D. Burns (bruce.burns@sydney.edu.au)
School of Psychology, University of Sydney, Brennan MacCallum Bldg, A18, Sydney, NSW 2006, Australia

Regina Vollmeyer (r.vollmeyer@paed.psych.uni-frankfurt.de)
Institute of Psychology, Goethe University, Theodor-W.-Adorno-Platz 6, D-60323 Frankfurt/Main, Germany

Ulrich Kortenkamp (ukortenk@uni-potsdam.de)
Institute of Mathematics, University Potsdam, Am Neuen Palais 10, D-14469 Potsdam, Germany

Abstract

The three-space theory of problem solving predicts that the quality of a learner's model and the goal specificity of a task interact on knowledge acquisition: Learners having a good model should learn more with a nonspecific than a specific goal, which should not apply to learners having a poor model. This study tested this prediction using a computer based learning task on torques. Participants (N = 77 psychology students) either had to test hypotheses with a simulation of a lever system (nonspecific goal), or to produce given values for variables in this simulation (specific goal). In the good model condition but not in the poor model condition they saw the torque depicted as an area. Results revealed the predicted interaction. A nonspecific goal only resulted in better learning when a good model of torques was provided but not with a poor model. Our findings support the three-space theory. They emphasize the importance of understanding in studying problem solving and stress the need to study underlying processes.

Keywords: goal specificity, problem solving, three-space theory, scientific discovery learning

Introduction

Explaining Goal Specificity Effects

Multiple studies with a variety of tasks have demonstrated that people learn better when they work on tasks with nonspecific goals than on tasks with specific goals (e.g., Ayres, 1993; Geddes & Stevenson, 1997; Paas, Camp, & Rikers, 2001; Sweller & Levine, 1982; Vollmeyer & Burns, 2002). For example, in Sweller and Levine's finger maze task the blindfolded participants either had one finger at the finish point (i.e., specific goal) or had no information about the location of the finish point (i.e., nonspecific goal). Those with the nonspecific goal performed better. In this paper we will try to clarify the mechanism underlying goal specificity effects and in doing so explain and test a theory of when nonspecific goals will and will not help learning. In particular this theory emphasizes the role of the participants' general understanding of the task, which we refer to as their model.

A possible mechanism for the goal specificity effect draws on Cognitive Load Theory (Sweller, 1988; Sweller, Ayres, & Kalyuga, 2011). Sweller proposed that when a specific goal is given people tend to use means-ends strategies to solve the task; that is, they try to reduce the difference between the current state and the goal state. Thus, they have to keep in memory a lot of information, such as the goal state, the actual state, the relation between these states, and potential sub goals, which leads to high cognitive load (Ayres & Sweller, 1990; Owen & Sweller, 1985; Wirth, Künsting, & Leutner, 2009). As a consequence a reduced amount of working memory is available for learning through schema construction or concept development. In contrast, in tasks with nonspecific goals people do not need to make comparisons with a given goal state. Therefore cognitive load is lower, and thus capacity for learning is increased.

Our own explanation for the mechanism by which goal specificity affects learning emphasizes that the nature of the goal alters the strategy learners take. This perspective comes from dual-space theories of problem solving (Klahr & Dunbar, 1988; Simon & Lea, 1974), which describe problem solving as search of two interacting problem spaces: experiment space/instance space and hypothesis space/rule space. Experiment space contains all possible experiments that can be conducted within a task, that is, transformations of the task elements. Hypothesis space consists of all possible hypotheses or rules about the task. These can be tested by running experiments (i.e., movement in experiment space) and as a result of experimenting hypotheses can be confirmed or rejected and rules can be derived (i.e., movement in hypothesis space). From dual-space theories it can be inferred that tasks with specific goals can be solved by moving in experiment space whereas nonspecific goals may encourage an additional search of hypothesis space. Indeed, it has been shown that nonspecific goals induce a search of hypothesis space in terms of hypothesis testing (Burns & Vollmeyer, 2002; Künsting, Wirth, & Paas, 2011). Learners who do not focus on reaching a given goal in experiment space are more likely to explore a task thoroughly by testing hypotheses, and this can explain their better learning results.

A series of experiments have tested goal specificity predictions derived from dual-space theories (Burns &

Vollmeyer, 2002; Osman & Heyes, 2005; Vollmeyer, Burns, & Holyoak, 1996). Learners had to control a computer simulation of a linear system (e.g., biology lab or water tank) in which they could manipulate input variables that affect output variables. The links between those variables were unknown. The specific goal was to bring the system to specific values whereas the nonspecific goal gave them no values to reach but instead encouraged them to find the rules underlying the system. Learners with a nonspecific goal acquired a better knowledge of the system's structure than learners with a specific goal. Similarly, giving learners a hypothesis to test led to better learning compared to learners who had the same amount of information but were not induced to test a hypothesis (Vollmeyer & Burns, 1996). These results can be interpreted in line with dual-space theories, supposing that nonspecific goals or hypothesis instruction induce search of hypothesis space and thus lead to better learning. So the mechanism through which a nonspecific goal improves learning is its encouragement of search of hypothesis space.

A Three-Space Theory

The theoretical perspective of a nonspecific goal as encouraging search of hypothesis space led to an important question about goal specificity: Should hypothesis testing always produce better learning results? Indeed, recent studies suggest that the goal specificity effect is not uniformly found and might be reversed under certain conditions (Pretz & Zimmerman, 2009; Zanga, Richard, & Tijus, 2004). There may be situations in which search of hypothesis space is unsuccessful, thus nonspecific goals may not always facilitate learning. This could be the case for someone who starts with a hypothesis space that is limited to inappropriate hypotheses. To deal with this Burns and Vollmeyer (2000) suggested the theoretical framework of a three-space theory. This extended dual-space theories by proposing a third space, model space, which contains possible models of a task or a domain.

Empirical evidence that led to the assumption of a model space came from studies of learning about linear systems. Burns and Vollmeyer (2002) found that some participants considered that there might be interactions between variables which was not the case. Thus, participants seemed to hold a certain model of linear systems which determined the hypotheses they took into account and thus defined their hypothesis space.

The three-space theory assumes that model space determines hypothesis space just as hypothesis space determines the appropriate experiment space. Further we assume that for any task a learner always has some model of how it might work, but the quality of that model can vary. The current state in model space constrains hypothesis space and determines the hypotheses that are considered plausible to test, so a good model is one that provides a searchable hypothesis space containing the appropriate hypotheses. Thus, when search of hypothesis space is encouraged, either through a nonspecific goal or hypothesis instruction, an appropriate model provides a searchable hypothesis space. In contrast, an inappropriate (i.e., incorrect or incomplete) model can define a set of inapplicable hypotheses to test or simply a set of hypotheses too large to search effectively, and thus may be misleading and actually hinder learning.

Our Hypothesis

An implication of the three-space theory is that we would expect an interaction between the quality of a learner's model and goal specificity, if the variation in model quality is great enough to have an impact on performance. When participants have a good model then their hypothesis space is searchable and thus encouragement to do so via a nonspecific goal should result in better learning than a specific goal, just as we have found before. However, when learners have a poor model then encouraging search of a poorly defined hypothesis space (via a nonspecific goal) should lead to little learning. Instead a learner with a poor model may be better off focusing on search of experiment space, which a specific goal would encourage. This is possible because a focused search of experiment space may yield more knowledge than would a haphazard search of experiment space produced by attempts to test the wrong hypotheses. So we predict an interaction between a manipulation of goal specificity and the manipulated quality of a learner's model. Whether learners with a poor model would actually learn less with a nonspecific goal than a specific goal is hard to predict, because it may depend on characteristics of the task. For example, a task, in which simply pushing towards the goal helps performance, might benefit more from a specific goal than a nonspecific goal when the model is poor. It is also possible that if careful testing of experiment space could help formation of a better model, then such movement in model space would be more likely with a specific than a nonspecific goal. Confirmation of such an interaction prediction would support the three-space theory and have implications for when different goals lead people to learn most effectively.

To test our interaction hypothesis we used a task in which participants manipulate a simulated system of levers and forces in order to learn about torques. Kistner, Burns, Vollmeyer, and Kortenkamp (2014), using a similar task, found that people had different levels of understanding of how such systems work, so such a task should be amenable to manipulation of a participant's model quality. In addition we found that we could manipulate goal specificity by either giving participants a set of questions to answer (specific goal) or asking them to test hypotheses (nonspecific goal).

We tested the interaction hypothesis, which in terms of this task was that on a posttest there would be an interaction between a manipulation of model quality and goal specificity such that participants given a good model would learn more with a nonspecific than a specific goal whereas those given a poor model would not be affected by goal specificity.

Method

Participants

Participants in the study were 99 first year psychology students at the University of Sydney who took part for partial course credit. All of the participants had studied some physics at high school (64% as part of compulsory science classes up to Year 10, 24% chose to study it in the last two years of high school, 12% at university). In this experiment participants were supposed to work with an unfamiliar task to acquire new knowledge in a physics domain, so those with high initial knowledge could not be expected to gain much from working with our task. Therefore we could not test the impact of our manipulations on their posttest level. For this reason participants who scored high in a pretest (described below) were identified. *High* was defined as one standard deviation above the mean, so the 22% scoring more than 3 out of 7 points were excluded from further analyses. The resulting sample of 77 participants had a mean age of 19.94 (*SD* = 4.65) and 74% were female.

The Computer Simulation

Participants worked with a computer simulation of a lever system (see Figure 1), which was created with the interactive geometry software "Cinderella" (Richter-Gebert & Kortenkamp, 1999). The simulated lever system consists of two lever arms (l1 and l2) on both sides of a fulcrum (A) and two forces (F1 and F2). The left side of the lever system, consisting of Lever Arm l1 and Force F1, is shown as inside a torque meter (grey box). Lever Arm l1 is fixed to be 8m in length and the Force F1 adjusts automatically in order to keep the lever system balanced as features on the right side of the lever system are manipulated. Outside the grey box the lever system continues with Lever Arm l2 and Force F2.

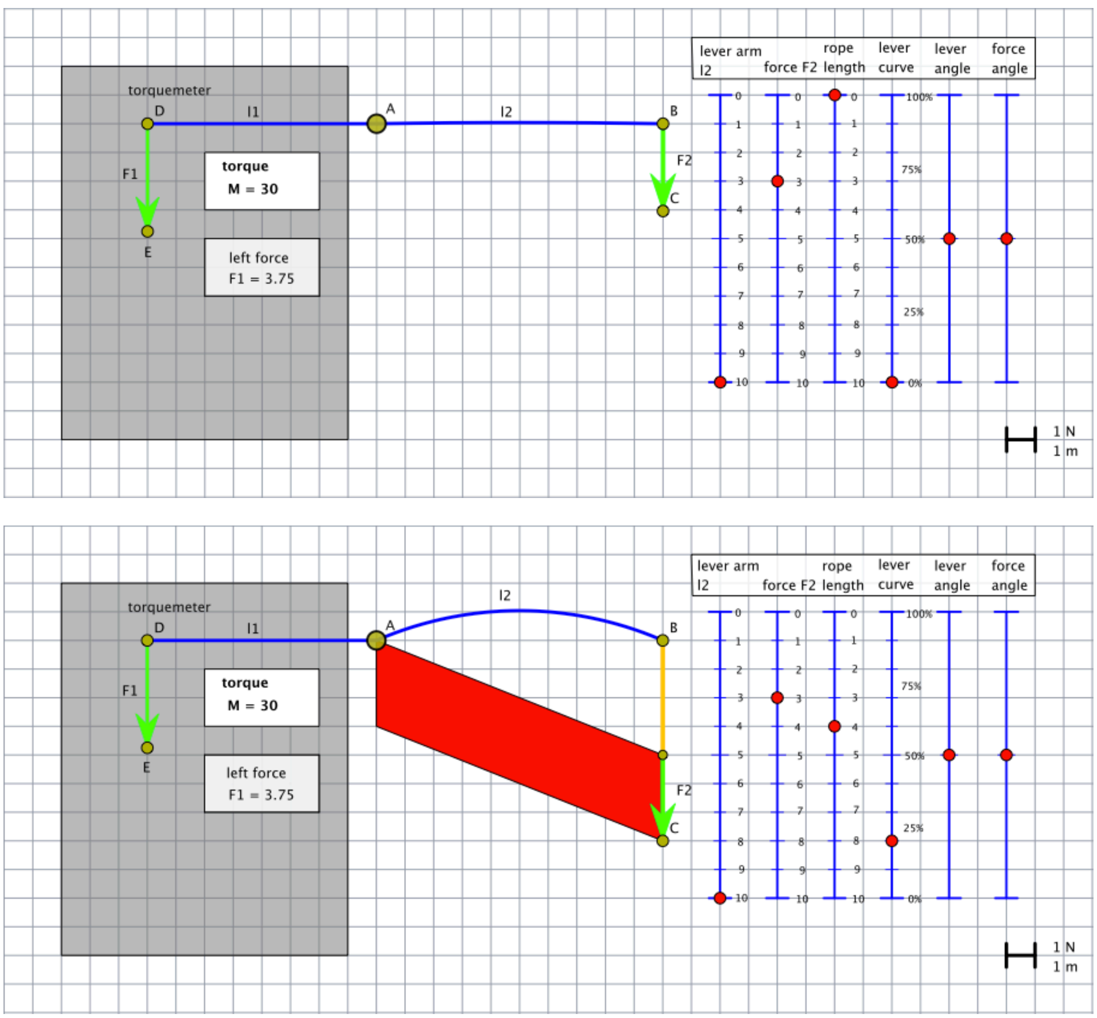

Figure 1. Illustration of the computer simulation of a lever system. The upper panel shows the simulation in the poor model condition. The basic setting of the simulation is shown. The lower panel shows the simulation in the good model condition. Lever curve and rope length have been manipulated in this setting.

The characteristics of Lever Arm l2 and Force F2 can be manipulated by six controllers next to the lever system. They allow for adjusting: (1) The length of Lever Arm l2: Lever Arm l2 can be made longer or shorter. (2) The magnitude of Force F2: Force F2, depicted by the arrow, can be made stronger or weaker, which is represented by increasing or decreasing the length of the arrow. (3) The length of a rope that can be fixed at Point B: By using this controller it is possible to integrate a rope into the lever system that applies at Point B. Force F2 then is no longer applied at Point B, but instead at the end of the rope (see Figure 1). (4) The degree of the lever curve: Lever Arm l2 can be curved by using this controller (see Figure 1). (5) The angle of Lever Arm l2 in relation to Lever Arm L1: In both panels of Figure 1 l2 is horizontal with l1 but it can also be angled upwards or downwards. (6) The angle of Force F2 in relation to Lever Arm l2: In both panels of Figure 1 F2 pulls in a downward vertical orientation but it can also pull in different directions.

When doing manipulations by using the controllers, the values for torque and Force F1 shown in the torque meter adjust. Thus, participants can observe the effects of their manipulations. By working with this simulation participants could acquire some understanding of the domain of torques and learn about the variables that determine torques.

Research Design

The study followed a 2 (goal specificity: specific goal [SG] vs. nonspecific goal [NSG]) x 2 (model quality: good vs. poor) design. Participants were randomly assigned to one of the four conditions: nonspecific goal with good model (NSG/good, $n = 20$); nonspecific goal with poor model (NSG/poor, $n = 21$); specific goal with good model (SG/good, $n = 15$); and specific goal with poor model (SG/poor, $n = 21$).

Goal specificity manipulation To vary goal specificity participants received two different task assignments. The aim of the SG condition was to induce search of experiment space, so in this condition participants received seven tasks incorporating 16 subtasks on paper sheets. For each task, participants first had to adjust the variables in the computer simulation according to a given basic setting. Then, they had to manipulate the simulation in a specific way, for example, to produce certain values for given variables, to read resulting values for other variables and to write them down in tables provided on their sheets. An example task was "Try to adjust force F1 in the torque meter to a value of 5 by varying lever arm l2 and force F2. Then, try to adjust it to a value of 7. In each case read the approximate values for l2 and F2 and enter them in the table below." The seven tasks were chosen in a way that they covered all relevant aspects that could be discovered with the simulation.

The aim of the NSG was to induce search of hypothesis space. The participants' task was to formulate and write down hypotheses about relationships between the variables in the computer simulation and to test them using the simulation. Therefore they got an overview of all variables and a short introduction on how to formulate the hypotheses. Participants wrote down their hypotheses on the provided sheets of paper. After testing each hypothesis they could mark whether it was confirmed, disproved, or needed further investigation.

Model quality manipulation To implement variation in model quality participants worked with two different versions of the computer simulation. From Kistner et al. (2014) we knew that a good model of torques in the context of similar simulations is conceptualizing torque as the area of the parallelogram spanned by force and lever arm. Thus, in the good model condition this parallelogram was depicted in the simulation as a red area, which adjusted when variables were manipulated (see Figure 1). So, participants could directly observe how the torque was affected when they worked with the simulation. Furthermore, when being introduced to using the simulation participants were informed that the red area was equal to the torque and that for a lever to be balanced the torques on both lever arms must be equal. This information was missing in the poor model condition, which also did not depict the area of the parallelogram spanned by force and lever arm (see Figure 1).

Assessment Instruments

Knowledge tests A pretest similar to the one in Kistner et al. (2014) was used. This contained four items on factual knowledge about torques (Cronbach's $\alpha = .63$). Examples are participants being asked to state the meaning of the term torque and being asked to compute the torques of a given lever system. Participants could score up to seven points in the pretest. The 17 items of the posttest (Cronbach's $\alpha = .80$) were of different formats. In addition to the four items of the pretest it included 11 multiple choice items that presented a specific setting of the computer simulation shown in a picture above. Every item began with a prediction, for example, "If force F2 increases, then …", and participants could choose from among four statements to complete the prediction. In another multiple choice item participants had to choose the correct formula(s) for the lever rule. Finally, they were asked to state a formula for calculating torques. The posttest had a maximum score of 33 points.

Manipulation check A manipulation check tested whether participants in the good model condition adopted the intended model of torque as the area of the parallelogram spanned by force and lever arm. Therefore, participants were given two figures like Figure 1 (without the red area) with the controllers set a certain way, and for each they were asked to do three tasks: (1) to draw the torque area onto the figure, (2) to compute the torque area, and (3) to state the torque magnitude. A maximum of nine points could be scored (across all components, Cronbach's $\alpha = .87$).

Procedure

Participants began by completing the pretest. Then they read short introductions to important terms in the context of torques (i.e., force, lever, and torque) and were shown a graphic of the simulation with an explanation of how to use it. Participants then had 30 minutes to work with the simulation according to their condition. Participants could not proceed until this time was over and they were prompted to continue working until the end. Afterwards, they filled out the posttest and completed the manipulation check. Altogether the procedure took about 60 minutes.

Results

Manipulation Check

To check whether our model quality manipulation was effective we first examined the manipulation check. Participants in the good model groups obtained significantly higher scores ($M = 2.54$, $SD = 3.21$) for drawing and computing torque parallelograms than participants in the poor model groups ($M = 0.43$, $SD = 1.55$), $F(1,75) = 14.26$, $MSE = 5.99$, $p < .001$, $\eta^2 = .16$. This is evidence that the manipulation indeed provided participants in the good model group with the model of torques as the area of a parallelogram.

Testing our Hypothesis: Interaction Between Goal Specificity and Model Quality

Based on the three-space theory we predicted that the effect of goal specificity would interact with model quality such that the difference in performance in the posttest between the good model condition and the poor model condition would be higher with the NSG than with the SG.

Table 1: Descriptive statistics of pre- and posttest for each of the four groups.

	NSG/ good *M (SD)*	NSG/ poor *M (SD)*	SG/ good *M (SD)*	SG/ poor *M (SD)*
Pretest	1.20 (1.11)	0.71 (0.90)	0.87 (1.06)	0.57 (0.87)
Posttest	16.85 (6.12)	10.33 (4.22)	12.93 (2.94)	12.43 (4.01)

Table 1 shows the results of the four groups in the pre- and the posttest. The groups did not differ in their pretest scores (no main effects, no interaction effect, all *p*s > .05). Figure 2 illustrates for the posttest the significant interaction found with a two-factorial ANOVA between goal specificity and model quality, $F(1,73) = 8.24$, $MSE = 20.70$, $p = .005$, $\eta^2 = .10$. Participants with a good model performed better in the posttest when they worked with a NSG compared to a SG, $F(1,33) = 5.21$, $MSE = 25.26$, $p = .03$, $\eta^2 = .14$. For participants in the poor model conditions goal specificity did not make a difference with regard to the posttest, $F(1,40) = 2.72$, $MSE = 16.95$, $p = .11$. Looked at another way, for participants given a NSG model quality had a large impact on the posttest, $F(1,39) = 15.87$, $MSE = 27.42$, $p = .001$, $\eta^2 = .29$, whereas for those given a SG model quality was irrelevant, $F(1,34) = 0.17$, $MSE = 13.00$, $p = .68$.

Thus, the results were in line with our hypothesis of an interaction between goal specificity and model quality: encouragement to search hypothesis space (via a NSG) only appeared to help learning when the learner's model was good enough for the hypothesis space to be relatively easily searchable.

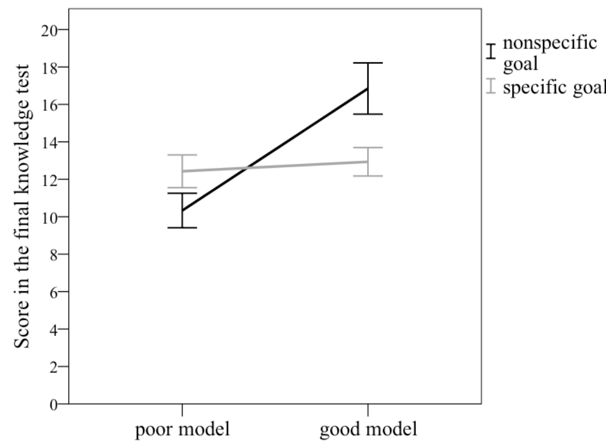

Figure 2. Mean posttest scores for each of the four groups with standard error bars.

Discussion

We started with the question of do nonspecific goals always help learning? We found that participants with a good model differed in their knowledge acquisition depending on their goal specificity, whereas goal specificity played little role for participants learning with a poor model. A good model should help nonspecific goal learners to develop more and better suited hypotheses and therefore they acquire more knowledge when encouraged by a nonspecific goal to test hypotheses. However, with a poor model a nonspecific goal might even hinder knowledge acquisition as nonspecific goal learners can get stuck with inappropriate hypotheses. Such a negative effect of a nonspecific goal was evident in the mean knowledge scores but was not statistically significant.

Our results are in line with the three-space theory from which we derived the hypothesis on the interaction between the experiment, hypothesis and model spaces. In this theory hypothesis testing (i.e., search in hypothesis space) is not always more advantageous than pure experimenting (i.e., search in experiment space), it depends on model quality. Moreover, this study emphasizes that it is not manipulating goals per se (i.e., goal specificity) that is responsible for learning; instead it depends on the underlying processes.

A limitation of the study was that we examined only one outcome variable, which was the score in a knowledge test. From the expected difference in knowledge we conclude that learners with a good model and a nonspecific goal had more effective hypothesis testing than learners with a poor model and a nonspecific goal. However, we had no direct indicator of the learning processes that we were postulating. Future research using this task should include mediating indicators for the effect of goal specificity and model quality on learning outcome.

If generalizable, the model quality by goal specificity interaction has practical implications for learning. Learners will always have some model of any task they are given, no matter how impoverished it is. Such initial models can be expected to vary as they will depend on a learner's prior knowledge. The model by goal specificity interaction suggests that prior knowledge may interact with other manipulations if those manipulations affect how the learner approaches the task. Therefore, the same intervention could improve learning for one person and be detrimental for another. Thus, the learners' prior knowledge needs to be taken into account when deciding how best to design a learning task.

Acknowledgements

This research was financially supported by the Deutsche Forschungsgemeinschaft (VO 514/17-1).

References

Ayres, P. L. (1993). Why goal-free problems can facilitate learning. *Contemporary Educational Psychology*, *18*, 376–381.

Ayres, P., & Sweller, J. (1990). Locus of difficulty in multistage mathematics problems. *The American Journal of Psychology*, *103*(2), 167–193.

Burns, B. D., & Vollmeyer, R. (2000). Problem solving: Phenomena in search of a thesis. In L. Gleitman & A. K. Joshi (Eds.), *Proceedings of the twenty-second annual meeting of the cognitive science society* (pp. 627–632). Hillsdale, NJ: Lawrence Erlbaum.

Burns, B. D., & Vollmeyer, R. (2002). Goal specificity effects on hypothesis testing in problem solving. *The Quarterly Journal of Experimental Psychology*, *55A*(1), 241–261.

Geddes, B. W., & Stevenson, R. J. (1997). Explicit learning of a dynamic system with a non-salient pattern. *The Quarterly Journal of Experimental Psychology*, *50A*(4), 742–765.

Kistner, S., Burns, B. D., Vollmeyer, R., & Kortenkamp, U. (2014). An explorative study of search of model space in problem solving. *Journal of Cognitive Psychology, 26* (7), 818-829.

Klahr, D., & Dunbar, K. (1988). Dual space search during scientific reasoning. *Cognitive Science*, *12*(1), 1–48.

Künsting, J., Wirth, J., & Paas, F. (2011). The goal specificity effect on strategy use and instructional efficiency during computer-based scientific discovery learning. *Computers & Education*, *56*(3), 668–679.

Osman, M., & Heyes, C. (2005). Practice doesn´t always make perfect: Goal induced decrements in the accuracy of action- and observation-based problem solving. In B. G. Bara, L. Barsalou, & M. Bucciarelli (Eds.), *Proceedings of the twenty-seventh annual conference of the cognitive science society* (pp. 1690–1695). Mahwah, NJ: Lawrence Erlbaum.

Owen, E., & Sweller, J. (1985). What do students learn while solving mathematics problems? *Journal of Educational Psychology*, *77*(3), 272–284.

Paas, F., Camp, G., & Rikers, R. (2001). Instructional compensation for age-related cognitive declines: Effects of goal specificity in maze learning. *Journal of Educational Psychology*, *93*(1), 181–186.

Pretz, J. E., & Zimmerman, C. (2009). When the goal gets in the way: The interaction of goal specificity and task difficulty. *Thinking & Reasoning*, *15*(4), 405–430.

Richter-Gebert, J. & Kortenkamp, U. H. (1999). *The interactive geometry software Cinderella.* Berlin: Springer.

Simon, H. A., & Lea, G. (1974). Problem solving and rule induction: A unified view. In L. W. Gregg (Ed.), *Knowledge and cognition*. Potomac, MD: Lawrence Erlbaum.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, *12*(2), 257–285.

Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory*. Dordrecht: Springer.

Sweller, J., & Levine, M. (1982). Effects goal specificity on means-ends analysis and learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *8*(5), 463–474.

Vollmeyer, R., & Burns, B. D. (1996). Hypotheseninstruktion und Zielspezifität: Bedingungen die das Erlernen und Kontrollieren eines komplexen Systems beeinflussen [Hypothesis instruction and goal specificity: Determinants of learning and controlling a complex system]. *Zeitschrift für Experimentelle Psychologie*, *43*(4), 657–683.

Vollmeyer, R., & Burns, B. D. (2002). Goal specificity and learning with a hypermedia program. *Experimental Psychology*, *49*(2), 98–108.

Vollmeyer, R., Burns, B. D., & Holyoak, K. J. (1996). The impact of goal specificity on strategy use and acquisition of problem structure. *Cognitive Science*, *20*, 75–100.

Wirth, J., Künsting, J., & Leutner, D. (2009). The impact of goal specificity and goal type on learning outcome and cognitive load. *Computers in Human Behavior*, *25*, 299–305.

Zanga, A., Richard, J.-F., & Tijus, C. (2004). Implicit learning in rule induction and problem solving. *Thinking & Reasoning*, *10*(1), 55–83.