Two Plus Three Is Five: Discovering Efficient Addition Strategies without Metacognition

Steven S. Hansen (sshansen@stanford.edu)
Cameron McKenzie (crlmck@stanford.edu)
James L. McClelland (mcclelland@stanford.edu)
Department of Psychology
Stanford University, Stanford, CA 94305 USA

Abstract

When learning addition, children appear to perform a remarkable feat: as they practice counting out sums on their fingers, they discover more efficient strategies while avoiding conceptually flawed procedures. Existing models that seek to explain how children discover good strategies while avoiding bad ones postulate metacognitive filters that reject faulty strategies. However, this leaves unexplained how the domain-specific knowledge required to evaluate a strategy might be acquired prior to addition being mastered. We introduce a biased exploration model, which demonstrates that new addition strategies can be discovered without invoking metacognitive filtering. This model provides a fit to data comparable to previous models, with the considerable advantage of avoiding an appeal to knowledge whose source is not itself explained. Specifically, we fit the pattern of changes in strategy use over time as children learn addition, as well as the overall error rate and error types reported empirically. The model suggests that the critical element allowing strategy discovery may be prior learning, rather than metacognitive strategy evaluation. We close by offering several empirical predictions and propose that what others have called strategies might often be decomposable into elements that can be assembled on the fly as problem solving unfolds in real time.

Keywords: Mathematical Cognition; Strategy Discovery; Reinforcement Learning; Metacognition.

Introduction

Single-digit addition is one of the first hurdles children master on their way to mathematical competence. Given the importance of mathematics to educational attainment, it is unsurprising that the process by which children learn addition has received considerable attention (eg, Siegler & Jenkins, 1989). A remarkable observation from these studies is that, once they are equipped with the ability to count out sums on their fingers, children spontaneously (without instruction) exhibit faster strategies. Despite this willingness to innovate, children rarely arrive at a strategy that, when executed correctly, leads to the wrong answer. This poses a real problem for trial and error theories of learning. As they acquire new, faster strategies, how do children know which strategies to avoid?

Several attempts have been made to model the evolution of children's approaches to solving simple addition problems. The apparent absence of explicit instruction in the use of particular observed strategies would normally make reinforcement learning a candidate mechanism, and indeed several early models did have this character (Neches, 1987). However, the paucity of solution paths that involved faulty strategies appear to rule out the 'take a random step' style exploration used by most reinforcement learning models (Crawley, Shrager, & Siegler, 1997). Trial and error accounts were thus rejected, and replaced by a theory that posited a metacognitive mechanism with explicit, domain-specific content knowledge to filter out flawed strategy proposals. This mechanism allows the discovery of new strategies while producing only reasonable errors. However, it remains unclear how children could acquire the complex knowledge required to judge the appropriateness of their own strategy proposals. The acquisition of the metacognitive filter is neither explained nor explicitly modeled.

The approach taken in this paper is to circumvent this difficulty by proposing that a metacognitive filter may not be necessary in the first place. We accomplish this by modifying a standard trial and error, reinforcement-learning-based paradigm to be biased towards previously learnt actions. We note that children learning the addition task have already learnt to count out numbers of objects, count on their fingers, and perform addition using a finger-counting strategy (Siegler & Jenkins, 1989). As we shall demonstrate, instantiating a model with biases towards these actions obviates the need for a metacognitive filter. We also expand the notion of retrieval – a 'strategy' that circumvents the need to engage in a structured sequence of behaviors by simply recalling the correct final answer to a problem – by suggesting that retrieval might also occur for appropriate subparts of a larger problem. Our model makes several novel predictions about the discovery process and questions the notion that selection and discovery processes necessarily take place at the level of complete strategies.

Background

The data our model attempts to account for comes from a study examining 4 and 5 year olds' discoveries of new finger addition strategies (Siegler & Jenkins, 1989). Children are assumed to come to the task knowing a correct but inefficient strategy, and are observed in a series of sessions spread over approximately three months as they solve simple addition problems. Over this time, children gradually acquire strategies that lead to faster completion of the task.

In our view, it is important to frame the discovery process against the backdrop of relevant previously learnt procedures. The most important to our theory is what we will call the *count-list procedure* whereby the child learns to step through a stable ordering of the number words, sometimes while counting out fingers or other physical tokens. The count list is a prerequisite for learning addition and is known by all children in the study.

We also assume (following Davidson, Eng & Barner, 2012) children can perform the *how many* task, in which the child verbally goes through his count list in order, simultaneously pointing to the next in a set of physical tokens, then responding with the number reached when the items in the set have been exhausted. It is generally accepted that this behavior is well learnt by the time children are taught their first addition procedure. Finally, we assume children have mastered the *give-a-number* or *give-n* task, which involves providing a supply of tokens and asking the child to give the experimenter (or other target) a certain number of them. Children who can perform this task for numbers larger than 4 do so by stepping through the count list as they remove them one by one from the supply, stopping when they reach the requested number.

In the study we will be considering, children were enrolled in a preschool/kindergarten that taught a standard procedure, known as the 'sum' strategy, for correctly adding two numbers together. This procedure begins with the child counting from one up to the value of one of the two addends, while simultaneously putting up a finger or taking a token from a pile on each count. The remaining addend is counted out in the same manner. The child then proceeds to count off each finger/token that she has previously enumerated. For example, a protocol for the problem "2+3" might read: "one, two (while raising two fingers), one, two, three, (while raising three more fingers), one, two, three, four, five (while counting the raised fingers). Five." Crucial to our later modeling, this protocol can be reframed in terms of previously learnt procedures. The first step is a 'give-a-number' task where the number to be given is one of the addends. The second step is the same, but targeting the second addend. Finally, to produce the answer, a 'how many' task is performed on the fingers/tokens produced by the previous two tasks.

After prescreening sessions where the children's knowledge of the sum strategy was verified, the children were asked to solve addition problems across several sessions. The children predominantly used the 'sum' strategy at first, but adapted their procedures over time, generally moving to approaches that increased accuracy while decreasing time taken. The experimenters coded the children's behavior as falling into one of several discrete 'strategies' on a trial by trial basis. At no point was only a single strategy chosen for all problems. Instead, there were 'overlapping-waves' of strategies. As shown in Figure 1b, the distribution of strategies changed quite slowly, though in the last block of trials "challenge" problems were introduced (i.e. problem with one very large addend), which caused children who had already discovered the min strategy to expand its usage rapidly.

In the study, the majority of children discovered two new strategies, and generally did so in the same order. The first discovery was the *shortcut-sum* strategy, and this tended to occur very early on in the study.[1] This strategy involves counting up from one to the sum of the two numbers, though the interpretation of this behavior is a key question posed by our theory.

The second strategy, the *min* strategy, consists of counting from the larger of the two addends up to the sum. For example, for the problem "2+5", a possible protocol would be: "five, six, seven. The answer is seven." This strategy slowly gains dominance over both the shortcut-sum and sum strategies. While these transitions are occurring, children also gradually increase their reliance on the 'retrieval strategy', simply recalling the correct answer.

Given the categorical nature of their coding scheme, the researchers focused their analysis on when new strategies were discovered, how often they were used thereafter, and whether or not any incorrect strategies were ever used. The results of the study partially supported the idea that strategy change occurred through an exploration-based, incremental learning process. Children were not always able to describe or explain their new strategies to the experimenter. However, the authors also found no evidence that incorrect strategies were ever used and they argued that exploration of the space of possible strategies should lead to such errors. Though children did sometimes answer problems incorrectly, the authors argued that these errors did not represent the sort of conceptual mistakes one would assume children would make if they were randomly exploring the space of possible strategies.

The SCADS Model (Shrager and Siegler, 1998)

In the years following this study, Siegler and colleagues built several computational models to explain their data, culminating in SCADS (Strategy Choice And Discovery Simulation), which posits initial knowledge of the 'sum' strategy, a retrieval system for recalling answers based on associative learning, a module that proposes new strategies and another module that filters out proposals that do not meet criteria assuring their adequacy.

SCADS captures some aspects of the successive emergence of strategies shown in the behavioral data. However, the transitions in learning are far more rapid than in the empirical data, and no account is given for how children would acquire the posited metacognitive filtering mechanism. It is this gap that we attempt to address.

[1] These transitions are quite gradual on the time scale of the Siegler and Jenkins (1989) study, and some participants are already on their way into adopting these strategies at the outset of the study, but they are easily seen in aggregate data over longer time periods (Svenson & Sjöberg, 1983).

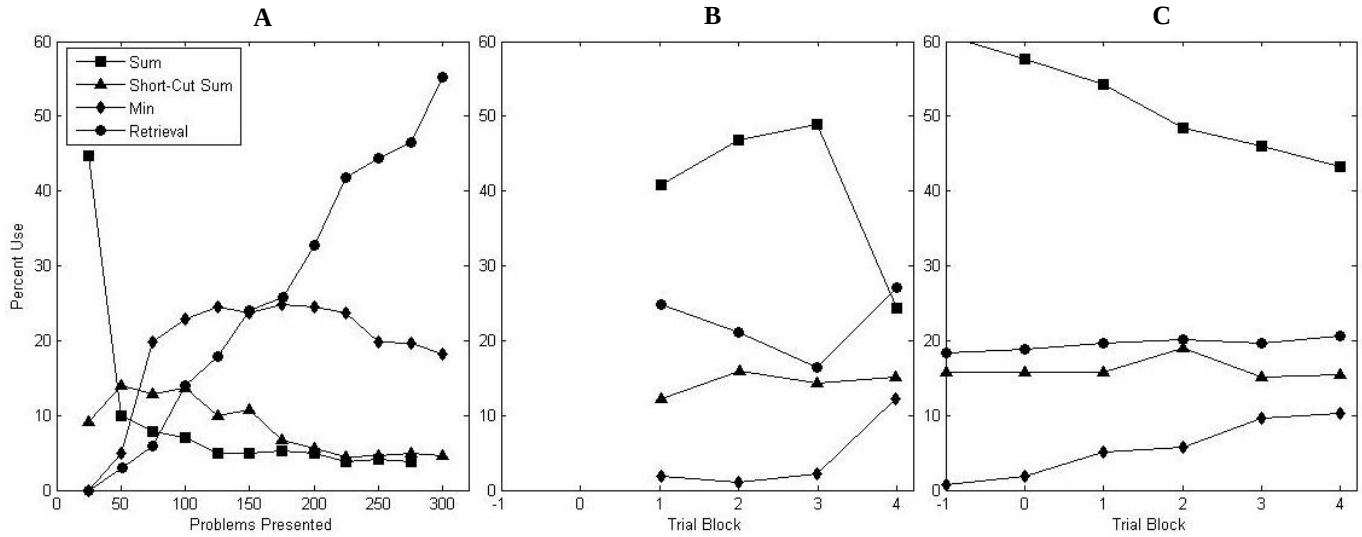

Figure 1: A) The SCADS model B) Behavioral strategy usage data across four blocks of 35 problems, average over 8 subjects. 60 prescreening trials were also administered but not recorded here. The abrupt change from the third to final block was due to a large intervention involving a spike in problem difficulty at the beginning of block 4. C) The biased exploration model, with data averaged over 100 runs. The milder strategy transitions are due to the lack of intervention

The Biased Exploration Model

Our model approaches the problem of strategy evolution through the use of a standard reinforcement learning system. It attempts to do away with the domain specific strategy proposal and filtering modules of SCADS. It avoids incorrect strategies because action is biased towards related, previously learned, procedures. The key insight arises by breaking down the two main strategy discoveries ('shortcut sum' and 'min') into the component steps needed to allow a new policy to arise from a predecessor.

For the 'shortcut sum' this means making two critical exploratory steps away from the existing 'sum' strategy policy. The first is to continue going through the count list after reaching the end of the first addend, rather than starting the count over at one for the second addend. The second is to stop going through the count list once the correct numeral is uttered. This second step can be seen as relying on problem specific knowledge, but avoiding reliance on recall by replacing it with an easier recognition problem whereby the child merely has to stop counting when the value reached feels like it is correct.

Exploration of this shortcut sum strategy can take place without assuming there is uniform exploration across all possible states and actions. We propose that the previously learnt counting procedure gives children a tendency to continue counting even when the first addend is reached. Thus, whilst the majority of the time the model chooses the sum strategy, occasionally a latent tendency to perform the related counting task takes over and an 'exploratory' step is made. Critically, this exploratory step speeds up task performance but does not lead to an error.

We should stress here that the discovery process proposed above is very different from prior proposals, which focus on realizing the redundancy in having to recount both addends (Shrager & Siegler, 1998; Neches, 1987). Which of these two conceptualizations is the better account of human behavior is an empirical question, and the two mechanisms are not necessarily mutually exclusive.

The size of the second addend is positively correlated with error rate when the shortcut-sum strategy is used, and this has been seen as evidence that the strategy was discovered to eliminate redundancy. It has been hypothesized that the increased error rate comes from the child having to hold two numbers in mind (one for the total count, and another for the count within the second addend). However, our recognition account is also compatible with this increased error rate, as larger second addends allow more chances to terminate the counting process, as well as providing more time to forget the problem, thus lowering the chance of recognizing the answer.

The second transition, from 'shortcut sum' to 'min' involves (a) skipping the counting out of the addend chosen to be dealt with first and (b) choosing the larger of the addends as the first one to deal with. Skipping the counting of an addend can be seen as a sort of retrieval, wherein it is the result of the subtask that is recalled rather than the answer to the entire problem. This subtask structure is part of the 'sum' strategy, which contains two instances of the preexisting 'give-a-number' task, first learned prior to encountering addition. Thus part (a) of the 'min' strategy comes about by starting with the 'shortcut sum' strategy, but instead of performing the full 'give-a-number' task for the first addend and then counting on, the child retrieves the end state of this subtask and then counts on. Part (b), choosing the larger of the two addends first, follows quickly from random exploration due to the inclusion of time cost in the reinforcement learning algorithm: The reward signal associated with producing the answer comes sooner in this

case, so the effective reward for dealing with the larger addend first is greater.

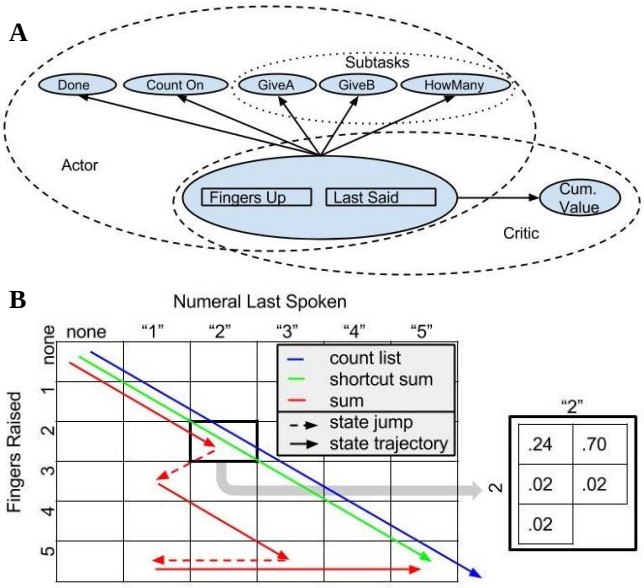

Figure 2: A) The actor-critic architecture as instantiated in the biased exploration model. The discrepancy between the critic's predicted cumulative value and reality is used to update both the critic and the actor. B) The state space of the biased exploration model for the problem 'two plus three', with arrows representing a sample trajectory for the sum (red) and shortcut sum (green) strategies as well as the count list procedure (blue). Each cell contains propensities for executing each of the five possible actions (see main text) at that state. The cell in the inset shows the propensities when in the 'two fingers up, just said two' state.

Implementation We implemented the model within the actor-critic reinforcement learning architecture. While this architecture has traditionally been chosen for its relative biological plausibility, here its utility comes from the fact that an actor-critic model learns by modifying the current policy. This feature limits how drastically a single learning step can affect the behavior of the agent (Sutton & Barto, 1998). This is important in part because it prevents the large number of errors typically associated with reinforcement learning. Since the initial policy (the explicitly taught sum strategy) is accurate, the model avoids a big change unless it consistently outperforms this existing solution.

As shown in Figure 2a, the actor-critic model consists of two main parts: the actor who selects actions to perform based on the current state, and the critic who predicts the expected cumulative reward of the actor at that state. At each time step an action is selected by treating the actor's action propensities as probabilities (via the softmax function). The action is performed, which modifies the state of the agent and produces some reward value. The critic is then able to see whether or not its prediction was better or worse than expected by comparing it against this actual reward plus its expectation at the new state. This signal, known as the TD error, is used to update both the critic and the actor.

All of the knowledge the model has about what actions to execute is stored in 10 two-dimensional tables, one for the state space of each of the 10 possible problems, as shown in Figure 2b. Each table can be imagined as a 6x6 square, where each cell represents a state of the world relevant to solving the specific problem associated with the table. The first dimension represents the number of fingers/tokens currently raised, from 0 to 5 (we only consider addition problems with sums up to 5, though nothing prevents the extension to larger problems). The second dimension is an echoic buffer that represents the last numeral uttered.

Each cell of the table contains 5 values, each representing the propensity towards taking a specific action in that state. There are 5 possible actions: perform a give-a-number subtask on the first addend, perform a give-a-number subtask on the second addend, perform a how-many subtask on the fingers currently raised, raise one more finger and say the next number in the count list, state that your previous utterance is your final answer. The first three actions are referred to as subtasks, since they involve a mediating process before affecting the state. When called upon to perform one of the subtask actions, either the end state is retrieved, or failing that, the whole subtask is carried out without interruption.

In addition to these actions, the agent tries to retrieve the sum at the start of each problem, and the action selection process described above only occurs when this initial retrieval fails. The assumptions made for retrieving this sum (as well as the subtask end states) are taken straight from SCADS. This memory mechanism, the 'distributions of associations' model, learns by accumulating association strengths between the task at hand and the various possible end states, with the idea being that it will converge to the correct answer as this is the most common end state for an agent competent at the task. When called upon to make a retrieval, a threshold is stochastically set and compared to each association strength. If no associations are higher than this threshold, then the retrieval fails to return an answer. If multiple associations are above the threshold, the retrieved association is randomly chosen from this set (Siegler & Shrager, 1984).

We give the model the initial 'sum strategy' policy by looking at all of the states encountered when executing this strategy, and setting the model's action preferences in these states to be consistent with the actions taken by the strategy. A weak preference for actions consistent with proceeding through the count list while putting up fingers is also encoded. As noted above, this is critical to the discovery of the shortcut-sum strategy in the model.

The model was trained on problems randomly sampled from cases where the sum was no greater than 5 and averaged over 100 runs. Since a tabular version of the Actor-Critic architecture was used, the top-level policy information (which subtask or primitive action to select) about each addition problem was learnt independently.

However, the give-a-number subtasks were shared across problems (for example, there was a single give-3 subtask, which could be carried out by iterating up to 3 or by producing 3 fingers all at once). While strategy learning may have a degree of problem specificity, as in the current model, sharing across specific problems seems likely, and some proposals on how to do so are put forward in the discussion section.

Results

The primary concern of this article was to demonstrate the viability of an exploration-based model of strategy discovery in addition. Following previous work (Shrager & Siegler, 1998), we focus on the qualitative fit to the pattern of strategy use as a function of problems encountered. A principal claim of the model is that avoidance of implausible errors can occur without metacognitive filters, so we also examine the model's errors, including the overall rate and types of errors. Whilst preferable, quantitative assessment of model fit is not possible, as detailed raw data is not available (Siegler, personal communication).

To ensure a valid comparison between our model and the available data, our model was trained for the same number of trials as the human participants (4 blocks of 35 trials). We trained the model for 2 preliminary blocks (labeled blocks -1 and 0 in Figure 1c) prior to this to account for the prescreening trials the participants received.

Strategy Distribution Dynamics Our model's strategy choices over time are shown in Figure 1c, where the number of correct trials is plotted for each strategy for each successive block of 35 trials. Since strategies are not explicitly represented anywhere in our model, the action sequence for each trial was examined to specify the strategy. We omit from figure 1 strategies which never achieved a usage rate greater than 5%. Additionally, the min strategy is discovered around the same time as in the study, which co-occurs with dropping usage of the sum strategy. In the study, there is an abrupt change from the third to final block; this was due to the inclusion of challenge problems (not yet addressed in our simulations) at the beginning of block 4.

Some of the model's solutions did not fit into one of the preexisting strategies, but played a significant role in the usage dynamics of the model. Specifically, a strategy emerged whereby the larger addend was retrieved (i.e. by recalling the end state of its give-a-number subtask), but then the rest of the solution followed the steps of the sum strategy. This 'retrieve larger, then sum' strategy (which occurred on 13% of trials averaged across the six blocks) played a crucial role in setting up the order dependence later needed in the min strategy. Having order dependence develop here solves the problem previous trial-and-error accounts have had where min discovery relied on first discovering the 'count from first' strategy, which is rarely used in children (Neches, 1987). While the 'retrieve larger then sum' strategy has not been reported in children to our knowledge, instances might have been lumped together with the sum strategy, based on their operational definition of the sum strategy as putting up fingers for each addend (agnostic as to whether they are simultaneously counted) and *then* counting them together. On this basis, we lump our model's data for the 'retrieve larger, then sum' strategy together with the prototypical sum strategy for comparison with the microgenetic study data.

Error Analysis Both the kind and quantity of the model's errors fell within the bounds of a typical child in Siegler's microgenetic study or a cross-sectional study covering similar addition problems (Siegler & Jenkins 89, Svenson & Sjöberg 83). The error rate averaged across 100 trials of each problem type was 13.4%, compared to 15% in the Siegler study. The problems Shrager and Siegler had assumed went hand in hand with trial and error learning, such as counting out a single addend twice, were also absent. This was determined by examining each erroneous trial and summing across those where the model made identical steps. Categorizing each unique error would be quite time prohibitive, but over 50% of errors occurred in a small number of unique action sequences. Of this group, the vast majority of the errors were in retrieval, with a failure to inhibit counting beyond the correct answer being the only other significant category of errors.

The retrieval errors have strong empirical support, and occur in our model precisely because its retrieval system is very closely based on the existing literature (Siegler & Shrager, 1984; Siegler & Shipley, 1995). The failures to inhibit counting, or 'count on' errors, are a consequence of our approach. Our theory relies on children taking steps in line with procedures related to, but different from, the task at hand, specifically the count-list procedure. Occasionally the model fails to stop counting upon reaching the sum, which is precisely what is happening in the count on errors. Whilst the exact frequency is not reported, Siegler and Jenkins (1989) themselves report that children occasionally count past the correct answer. One strong empirical prediction of our model is that these errors should occasionally occur and be indicative of a child in the early stage of learning, before the shortcut sum strategy is consistently accurate.

Discussion

We set out to explore how the problem of strategy discovery might be solved without a metacognitive filter, while avoiding a high error rate and approximating the pattern of change in strategy use observed in Siegler and Jenkins, 1989. The biased exploration model showed this to be possible and additionally demonstrated that strategies can be composed by assembling parts on the fly, rather than being selected as units at the start of the problem. This allows fine grained variations in strategies to be used, and predicts that such variation, such as the 'retrieve larger, then sum' strategy, will be seen in behavioral data.

Distinguishing our account from that of the SCADS model and investigating the extent to which metacognition plays a role in the discovery process will be another focus of our future work. One area where the models make distinct predictions is in the rationale behind the use of the shortcut sum strategy. The SCADS model claims that children track the total while also counting out the second addend, while our account relies on habitually counting on until the sum is recognized, avoiding the need to keep track of the second addend. This may be amenable to empirical exploration. Self-reports might also be used to differentiate these accounts, but we stress that we do not claim children do not eventually discover a rationale for their actions. Our claim is only that they need not do so before the actions themselves emerge.

Another area for future work will be to address the problem-specific representations of our current model and to explore the consequences of this for the model's predictions. Sharing information between problems might simply accelerate the learning process, but more fundamental changes are also possible. For example, sharing could increase certain errors due to confusion of one problem with another, which would change the pressures that lead to strategy discovery.

Another approach we are exploring is to let a neural network control the policy across all of the problems (in this case, the problem state would be represented as an input feature vector), as this could allow a more nuanced sharing of discovery information to emerge (it is possible to see at least some versions of a table-driven model as an alternative implementation of this neural-network based approach).

Going forward, we plan to extend our model to account for another stream of evidence that has previously been used to support the notion of metacognition: the recognition of never-before seen strategies. Children that have been shown the min strategy before discovering it still rate it as better than an incorrect strategy (Siegler & Crawley, 1994). While this has previously been taken as support for the proposed metacognitive filter, we suggest that the biased exploration model can account for this data as well by using the agent's value function to evaluate novel strategies. Such an extension is indicative of our overall goal with this model: to set up a new foundation for self-guided learning that will allow a rethinking of the role of metacognition in strategy discovery.

Acknowledgments

Thanks go out to the whole Parallel Distributed Processing Lab at Stanford, for their support throughout the research process and to Robert Siegler and the reviewers for useful comments.

References

Crawley, K., Shrager, J., & Siegler, R. S. (1997). Strategy discovery as a competitive negotiation between metacognitive and associative mechanisms. Developmental Review, 17(4), 462-489.

Davidson, K., Eng, K., & Barner, D. (2012). Does learning to count involve a semantic induction?. *Cognition*, *123*(1), 162-173.

Neches, R. (1987). Learning through incremental refinement of procedures. Production system models of learning and development, 163-219.

Shrager, J., & Siegler, R. S. (1998). SCADS: A model of children's strategy choices and strategy discoveries. Psychological Science, 9(5), 405-410.

Siegler, R. S., & Shrager, J. (1984). A model of strategic choice. Origins of Cognitive Skills. Hillsdale, NJ: Lawrence Erlbaum.

Siegler, R. S., & Jenkins, E. (1989). How children discover new strategies. Lawrence Erlbaum Associates, Inc.

Siegler, R. S., & Crawley, K. (1994). Constraints on learning in nonprivileged domains. Cognitive Psychology, 27(2), 194-226.

Siegler, R. S., & Shipley, C. (1995). Variation, selection, and cognitive change. Developing cognitive competence: New approaches to process modeling, 31-76.

Sutton, R. S., & Barto, A. G. (1998). Reinforcement learning: An introduction (Vol. 1, No. 1). Cambridge: MIT press.

Svenson, O., & Sjöberg, K. (1983). Evolution of cognitive processes for solving simple additions during the first three school years. Scandinavian journal of psychology, 24(1), 117-124.