A rational cascade from teaching explains strategy discovery in development: The case of early addition

Merrick Giles†, Andrew Perfors & Francis Mollica
University of Melbourne, Australia
†merrickdgiles@gmail.com

Abstract

Without being taught, children learning addition discover strategies that are faster and less error-prone than those in their repertoire. Leading explanations of strategy discovery propose metacognitive mechanisms where children iteratively rewrite their existing strategies. Departing from this tradition, we propose that children make discoveries through a combination of reconstructive imitation during teaching, followed by rational inference under basic performance pressures. We instantiate our proposal with a model of Bayesian program induction with a prior for algorithmic simplicity and a likelihood that first encodes a pressure of mimicry (teaching phase), and then a pressure for speed and/or accuracy (individual practice phase). Regardless of whether the pressure is to respond faster or more accurately, the model robustly captures the long-observed developmental trajectory of strategy discovery.

Keywords: program induction; arithmetic development; strategy discovery; learning

Introduction

Our mathematical knowledge allows us to achieve otherwise inconceivable feats, from studying the structure of our universe to budgeting for our groceries. It is therefore vital that we pass this knowledge on to future generations. Yet it is not necessary to wait for formal schooling to begin teaching math (Fuson, 2012). Indeed, the cognitive artefacts of mathematical knowledge – things like arabic numerals and number words or morphology – are so littered throughout a child's early experiences that they can pick up some of it without explicit instruction, just by trying to understand the world around them (Almoammer et al., 2013; Gopnik et al., 1999).

The earliest stages of mathematical understanding, like the acquisition of the count list and the cardinality principle, are well-understood as a combination of rational construction (Xu, 2019) and social learning through teaching. Early in development, social learning takes the form of mimicry (Heyes, 2018). Children's first introduction to the concept of number consists of imitating their teachers by reconstructing a memorised count list (Fuson, 2012). Once they know the count list, children are able to rationally construct the meanings for small cardinalities (1-4) and potentially infer the cardinal principle without formal instruction (Piantadosi et al., 2012; Wynn, 1990; cf. O'Shaughnessy et al., 2023). Unfortunately, the subsequent stage of numerical development, basic addition, has, so far, evaded a similar explanation for its trajectory. Our aim with this paper is to offer one.

As described below, the acquisition of basic addition typically follows a series of stages known as the SUM-to-MIN trajectory. Because this pattern has previously eluded capture by state-of-the-art learning models involving Bayesian program induction, recent theorists have instead focused their explanations on metacognitive rewriting processes (Rule et al., 2020). However, we demonstrate here that the trajectory can be explained by incorporating the social-learning process of reconstructing the SUM strategy, through imitation, into an ideal-learner model. We show that as children are taught to reconstruct the first step in the trajectory (SUM), evidence accumulates not just for SUM but *also* for other potential strategies that approximate it. This sets the groundwork for a subsequent learning cascade across the latter stages in the trajectory, as children update their strategies in response to performance pressures in combination with exposure to the sorts of math problems seen in a realistic learning environment.

The SUM-to-MIN trajectory

The first step in learning to add is for children with mastery of the counting routine to acquire a memorised addition algorithm called SUM in which they add two numbers a and b by first counting from zero to a on one hand, then from to zero to b on the other hand, and finally counting from zero to all fingers raised: the sum of a and b. Children pick up the SUM strategy by mimicking their teachers, and then independently discover new algorithms that are faster and more accurate than it. The discovery of better strategies follows a consistent trajectory known as the SUM-to-MIN transition (Ashcraft, 1982, 1992; Baroody, 1984, 1987; Cowan & Renton, 1996; Geary et al., 2004; Groen & Resnick, 1977; Houlihan & Ginsburg, 1981; Kaye et al., 1986; Siegler, 1987; Siegler & Jenkins, 1989; Svenson & Broquist, 1975; Svenson & Sjöberg, 1983). As shown in Table 1, the SUM-to-MIN transition follows four stages (putting aside retrieval; see Siegler & Jenkins, 1989).

Stage 1. SUM. As previously mentioned, in this stage children mimic their teachers by using the slow and error-prone SUM strategy. Because it is easy on memory and uses familiar count-from-zero routines, the SUM strategy is likely optimised for teaching (Rule et al., 2020).

Stage 2. SHORTCUT-SUM. Instead of counting both digits from zero, children count to a and then count from a to b. This strategy cuts both response time and error rates in half.

Stage 3. COUNT-FROM. This stage is marked by the discovery that it is not necessary to count from zero to a at

Strategy	Algorithm	Description ($a + b$)
SUM	count(0, ∀, count(0, b, count(0, a, C)))	Count from 0 to a, count from 0 to b, then count from 0 to everything counted so far.
SHORTCUT-SUM	count(∀, b, count(0, a, C))	Count from 0 to a, then count from a to b.
COUNT-FROM	count(a, b, C)	Count from a to b.
MIN	count(a, b, C) if compare(a, b) else count(b, a, C)	If $a > b$, then count from a to b. Otherwise, count from b to a.

Table 1: Principle procedural strategies observed across development as children learn basic addition, sorted by order of acquisition. The strategies' algorithms can be formalised as simple programs composed of counting fuctions, which children often perform on their fingers.

all, instead counting directly from a to b. Studies vary in how much they report use of COUNT-FROM in practice (cf. Baroody, 1987; Carpenter & Moser, 1984; Ostad, 1997; Siegler & Jenkins, 1989), likely because it is difficult to empirically distinguish from the fourth stage (MIN); the two stages make identical predictions in half of all cases. Use frequency aside, children must discover COUNT-FROM to implement MIN.

Stage 4. MIN. Finally, children learn to start counting from the largest number. This strategy is the most complex addition algorithm since it composes the standard counting functions with numerical comparison and conditional evaluation. It is also the optimal strategy, minimising response time and maximising procedural accuracy. As a result, the use of MIN persists even in adulthood (LeFevre et al., 1996).

Explanations of the SUM-to-MIN trajectory

The SUM-to-MIN transition (including the intermediary stage/s) has been explored extensively in theoretical accounts of strategy discovery (e.g., Ashcraft, 1992; Baroody & Tiilikainen, 2003; Crowley et al., 1997b; Siegler, 1996; Svenson & Broquist, 1975), including many influential computational approaches (Jones & Van Lehn, 1994b; Neches, 1987; Shrager & Siegler, 1998). The common defining feature shared by these approaches is an appeal to metacognitive rewriting: the idea is that strategy discovery proceeds as children iteratively re-structure their addition algorithms to become more efficient.

The most influential of these models is known as the Strategy Choices And Discoveries (SCADS) model of Shrager and Siegler, 1998. As shown in Figure 1, SCADS encodes the SUM strategy at the onset of learning and iteratively deletes redundant actions until recovering COUNT-FROM. Although there is some ambiguity in the description of SCADS, the model seems to turn COUNT-FROM into MIN by always swapping the the addends to whatever order that minimises time and error.

While SCADS does successfully capture the trajectory of discovery from SUM to SHORTCUT-SUM to MIN, the model encodes arbitrary and implausible actions into its starting SUM strategy (see Figure 1). For example, it first stores addends (a, b) in its memory, but immediately deletes them. The function of this store-and-delete process is so that later, after many action deletions, the inital storage can serve as the initial numeral from which to start the counting procedure for MIN.

A metacognitive rewriting explanation of the SUM-to-MIN trajectory leaves open explanatory gaps regarding both *when* in development discoveries should take place and *how quickly* metacognitive editing should occur once it begins. For example, SCADS requires only a few metacognitive rewriting instances to acquire MIN. In order for SCADS to match the empirical timeline, the discovery process begins after the model solves the equivalent of approximately 70 problems. Shrager and Siegler (1998) argue that this delay captures the practice required to free cognitive resources from managing strategy execution and allocate them toward metacognitive discovery. While plausible, the account fails to capture environmental influences on the acquisition trajectory. Importantly, many children discover MIN soon after being introduced to difficult, large-numeral problems (Baroody & Tiilikainen, 2003; Baroody, 1984, 1987; Siegler & Jenkins, 1989); whereas, SCADS will learn MIN regardless of the problem complexity. Further, the model appears to predict the opposite of the empirical pattern: solving these more difficult problems should require more cognitive resources to manage strategy execution (and thus have fewer resources available for metacognitive discovery), thus making it harder to discover MIN rather than easier as SCADS predicts.

We therefore take an alternative approach to explaining the trajectory of strategy discovery: one that highlights the learning environment rather than metacognitive rewriting. Most importantly, we incorporate the influence of teaching – i.e., the process of reconstructive imitation – into our model at the onset of learning. After imitation, the learner updates their strategies in response to the demands of a realistic learning environment modelled after Siegler and Jenkins (1989)'s micogenetic study. Our work follows the tradition of rational constructivism (Xu, 2019), conceptualising discovery as the creation of addition algorithms that maximise accuracy and minimise response time. Therefore, our account is consistent with theories that emphasise the role of rational and adaptive environmental influence in strategic choice behaviour (Lieder & Griffiths, 2017; Marewski & Schooler, 2011; Payne et al., 1988; Siegler, 1987). However, unlike past approaches (Crowley et al., 1997b; Shrager & Siegler, 1998), we show that the same adaptive performance pressures can drive strat-

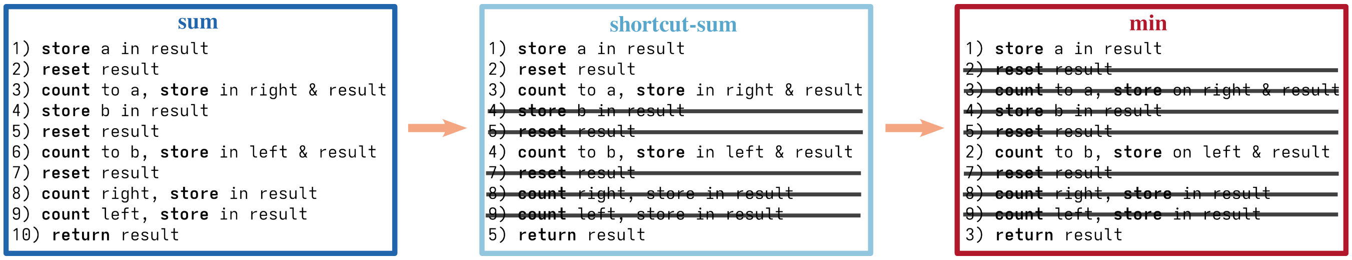

Figure 1: Metacognitive rewriting processes that give rise to arithmetic discovery in SCADS (Shrager & Siegler, 1998). The SUM strategy (left) is constructed to include the possibility for MIN via deletion. The model discovers SHORTCUT-SUM and MIN by deleting actions and swapping addend-orders to count from the larger addend, which is not represented in its algorithm.

egy discovery as well as strategy choice.

A Rational Cascade via Program Induction

Our model builds on the tradition of modelling developmental trajectories at the computational level using Bayesian program induction (Piantadosi, 2023; Piantadosi et al., 2012; Rule et al., 2020; Ullman & Tenenbaum, 2020). To define the model, we need to specify a space of hypothetical strategies a learner might consider, a prior over those hypotheses, a likelihood for the data under each hypothesis, and the environmental distribution of data observed by a learner. We can then simulate learning trajectories by analysing how the posterior over hypotheses changes as a learner observes more data. We instantiate our model in Python using LOTlib3 (Piantadosi, 2014).

Hypothesis Space. Following past work, we define our hypothesis space as the productions of a probabilistic context-free grammar. Our grammar includes numerical variables, the standard logic conditional if and the primitive functions known to children before addition learning in Siegler and Jenkins (1989)'s study: count and compare (Table 2). count(x, y, C) implements a noisy counting procedure starting from x, incrementing by one y times (with some error; see below) and returning the final count. Calls to count can be nested through the context variable C to capture strategies like SUM where the results of the nested calls are stored on "hands" in the context and the ∀ call increments through the stored values. compare(x, y) compares the magnitude of two numbers x and y and returns a boolean TRUE if $x > y$ and FALSE otherwise.

Prior. Following Goodman et al., 2008, we define the prior over a hypothesis as the production probability under the grammar that defines it, setting a uniform probability over grammar rules. Consequently, the prior imposes a pressure for algorithmic simplicity: strategies composed by fewer rules have higher probability. A simplicity prior such as this has been shown to provide a good explanation for adult concept learning (e.g., Feldman, 2000), developmental trajectories (e.g., Mollica & Piantadosi, 2022) and children'sexplanations for causal phenomena (Lombrozo, 2007).

Likelihood. Hypotheses specify generative functions resulting in an addition problem $a + b$ with a solution s and the response time RT taken to generate that solution. This yields a likelihood $P(s, RT|h)$. Because incrementing is as-

Primitives	Description
count(x, y, C)	Increment by 1 from x to y, return result
x if c else y	Do x if condition c is true, y otherwise
compare(x, y)	Return x if $x > y$, y otherwise
a, b	Variables for first and second addends
∀	Variable for everything counted so far
0, ..., 9	Variables for integers 0 to 9 (inclusive)

Table 2: Atomic actions used by the generative model to construct strategies; those with fewer actions have higher prior probability.

sociated with some probability of error (making the functions non-parametric), we use Monte Carlo methods to estimate the likelihood. For each problem that the learner sees, we use 100 simulations to inform equally weighted independent likelihoods for accuracy and response time, as described below.

Accuracy. We calculated the likelihood of a solution s given a specific hypothesis about the algorithm by assuming a fixed probability ϵ of making an increment error (either forgetting to advance or advancing twice). We set $\epsilon = 0.04$ to match the estimated errors-per-increment based on the empirical distribution of errors reported in Siegler and Jenkins (1989); however, the qualitative pattern of our results hold so long as accuracy is high enough to reconstruct the SUM strategy in teaching (i.e., $\epsilon \leq 0.06$). This suits empirical studies (Siegler & Jenkins, 1989), where children were required to reconstruct the sum strategy successfully for inclusion. Since children in Siegler & Jenkins (1989) were adept at numerical comparison at the onset of learning, we assume compare() functions are error-free. To ensure robustness, we applied additive smoothing ($\alpha = 0.01$) over the range of 0-100.

Response time. We assume a time cost $\beta_0 = 1$ for initiating each count or compare and a time cost $\beta_i = 1$ for each increment, as in previous work (Groen & Parkman, 1972). We then simulate response time by drawing from:

$$\text{RT}_{\text{sim}} \sim \mathcal{N}(\beta_0 z + \beta_i i, 10), \quad (1)$$

where z is the number of count/comparison initializations and i is the number of increments. We assume the response times are generated from a normal distribution with a mean equal to the average of the simulated RTs and standard deviation σ:

$$\text{RT} \sim \mathcal{N}(\widehat{RT_{\text{sim}}}, \sigma). \quad (2)$$

We use a wide standard deviation of 10 because of the variability inherent in the time taken to perform the strategy, and because we suppose children have only an approximate means of tracking the time it takes. The qualitative pattern of results is robust to any standard deviation value, only changing how many problems is necessary to produce strategy shifts.

Environmental Data Distribution. We follow the three phases in Siegler and Jenkins (1989)'s microgenetic study in order to simulate a typical learner's environment. The first phase (Teaching) captures how children acquire SUM by mimicking adults who teach it to them. The data consists of a series of math problems, which we simulate by randomly sampling addends uniformly between 1 and 5. Because the data is assumed to be generated by adults using the SUM strategy, all of the sums are always accurate, and the response times are distributed according to the SUM strategy.

In phases two (Easy), and three (Mixed), we aim to capture children's spontaneous discovery of strategies. The Easy phase presents problems with addends ≤ 5. The Mixed phase consists of 67% easy problems with addends ≤ 5, 26% problems with addends between 6 and 10, and 7% challenge problems, where one addend is ≤ 5 and the other is between 20 and 25. In these phases, we assume that children receive feedback about accuracy, and are motivated to be as efficient as possible (i.e., to minimise response time). Unlike most tasks, speed and accuracy in our model do not trade off. Faster strategies generally involve fewer increments, which also reduces the opportunity for counting errors.

Analysis Plan

Our goal is to demonstrate that we can recover the qualitative developmental trajectory of strategy discovery by appealing to rational learning based on environmental regularities instead of metacognitive rewriting. We thus simulate the posterior probability over the space of strategies h as an ideal learner who follows Bayes rule and observes data d (consisting of addends a and b, a solution s, and a response time RT):

$$P(h|d) \propto P(d|h) \cdot P(h) \tag{3}$$

As in past work (Goodman et al., 2008), we approximate the posterior using a Metropolis-Hastings sampler with tree regeneration proposals. Importantly, there is no metacognitive rewriting required or implicit in our search algorithm. Since this is an ideal learner model, we do not make any claims about how children search the space of hypotheses.

To ensure we found all of the high-probability hypotheses, we ran five chains initialised randomly plus five initialised at each strategy in the SUM-to-MIN transition. Each chain ran for 100,000 steps and was evaluated based on varying amounts of data (5, 10, 25, 50, or 100 problems) generated according to the three phases in the environmental distribution.

Among the top 1,000 hypotheses from each chain, we obtained 1,530 unique strategies. However, almost all of these strategies generated equivalent response (solution, RT) distributions. The largest equivalence class consisted of deviations from COUNT-FROM that included *false starts* (e.g., count(a, b, count(0, 3, C))). Like developmental psychologists, we do not consider these false starts as unique strategies and group them in with COUNT-FROM. After identifying all equivalent strategies, we reduced the space down to 170 strategies that generated unique response distributions. As most of the posterior mass is on the most likely a-priori strategy, we use the hypothesis with the greatest prior to stand-in for the equivalence class; these included all attested strategies in Table 1.

With few exceptions, the resulting hypotheses all reflect either a qualitative equivalence to the four strategies in Table 1, or an imperfect execution of them (with poorer accuracy or slower reaction time due to extraneous steps). Reassuringly, this suggests that these four strategies accurately represent the set of optimal addition strategies. It is also consistent with other ideal learning studies in which the high-posterior area occupies a very small region of the hypothesis space (Mollica & Piantadosi, 2022; Piantadosi et al., 2012).

Results

We find that the rational cascade implemented by our model is able to recover the SUM-to-MIN trajectory even without incorporating any metacognitive rewriting. Success is dependent on the initial Teaching phase in which learners are explicitly shown the SUM strategy, which guides subsequent discovery. Mirroring empirical studies, we find that the transition to MIN is facilitated by (but not dependent on) the presence of larger-addend problems in the Mixed phase. Interestingly, similar qualitative trajectories occur regardless of whether the pressure for efficiency is based on accuracy, speed, or both.

Capturing the SUM-to-MIN transition

Figure 2a shows the results of the simulation using the likelihood with both speed and accuracy. At the outset of learning, COUNT-FROM has the highest prior probability due to its advantage in algorithmic simplicity. However, its probability rapidly declines because it does not fit with observations in the Teaching phase (which favours SUM). Following the rise of SUM and the end of the Teaching phase (and thus the removal of the SUM response time pressure), the posterior of SHORTCUT-SUM, COUNT-FROM and MIN increase. At that point the strategy with the highest posterior is initially SHORTCUT-SUM: it has better accuracy and response time than SUM, but is still relatively consistent with the observed data from the Teaching phase. Subsequently, as the amount of speed-pressure data from the Easy and Mixed phases accumulates, the posterior density shifts first to the faster and more accurate COUNT-FROM strategy, until finally landing on MIN, which is the most complex but also the strategy with the optimal speed and accuracy. The model thus captures the empirical addition strategy trajectory.

Discovering MIN on small-addend problems

While the dataset of addition problems mirror that of Siegler and Jenkins (1989) and likely reflect the increasing difficulty of addition problems that children face in formal education, it

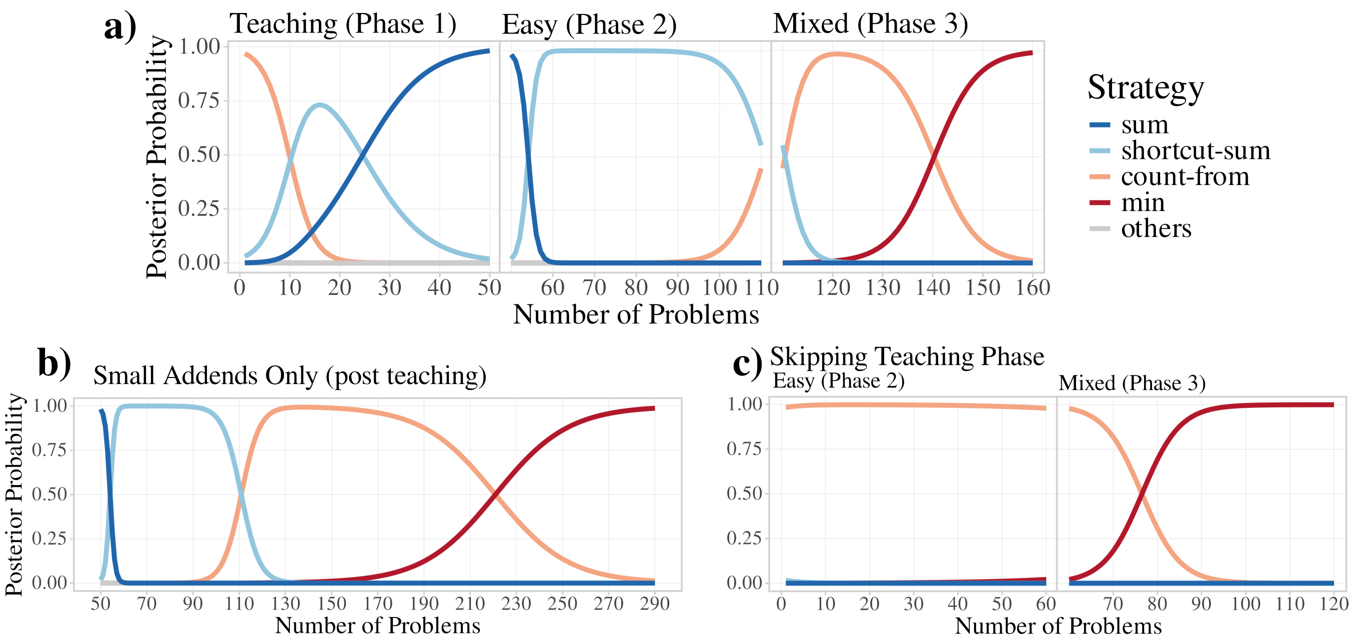

Figure 2: **Rational Cascade model simulations**. a): A standard simulation of the model with both accuracy and response time, on a realistic dataset that includes all three phases of Teaching, Easy problems, and Mixed problems. b): Simulation of the model where, post-Teaching, data only includes problems from the Easy phase. c): A simulation of the model but with no Teaching phase, where SUM is never favoured.

should be noted that some children in Siegler's study discovered MIN without the introduction of higher-addend problems (i.e., with exposure only to addends that ranged from 1 to 5). Simulations on our realistic three-phase dataset therefore leave open the question as to whether the model can capture the trajectory of children who discover min on Easy problems alone.

Figure 2b shows model simulations on small-addend problems, demonstrating that the model can in fact recover the trajectory even in this situation, and thus can explain the trajectory of children who do acquire MIN even without seeing large-addend problems. The model additionally predicts that it should take *longer* to generalise to MIN if one only sees Easy problems: on the Mixed dataset (Figure 2a), the posterior of MIN moves from ≈ 0 to ≈ 1 in 40 problems, whereas on the small-addend dataset (Figure 2b), over three times as much data (140 problems) is required. This is consistent with Siegler's (1989) study which found that children were slower to generalise to MIN based on small-addend problems, but generalised MIN more quickly when they saw Mixed problems.

Requiring training on SUM

In standard program induction models, a learning trajectory often falls out of the trade-off between the simplicity prior and the likelihood, which pull learners in opposing directions (e.g., Mollica & Piantadosi, 2022; Piantadosi et al., 2012). As Figure 2c shows, implementing our rational cascade model *without* the Teaching phase follows the same pattern, but fails to capture the empirical trajectory which starts at SUM. Instead, without Teaching, the strategy with the initially highest posterior density is the simplest COUNT-FROM strategy. As before, the superior likelihood score of the MIN strategy causes posterior density to shift to MIN with increasing amounts of data. It therefore produces a COUNT-FROM to MIN transition.

Qualitative equivalence of efficiency pressures

Remarkably, the SUM-to-MIN trajectory emerges regardless of whether speed or accuracy is emphasised. The trajectory from Figure 2a, which is based on a likelihood that incorporates both speed and accuracy, is qualitatively very similar to the trajectories derived from the Accuracy Only and Speed Only models shown in Figure 3. In the case of the Speed Only model (left panel), the only strong qualitative difference is that the posterior of the SUM strategy is raised faster during the Teaching phase. This is because the SUM strategy is no longer penalised for its relatively low accuracy (which it gets because there are more increments and thus more opportunity for error). In the Accuracy Only model (right panel), the strategy trajectory is broadly recovered but the transition to SHORTCUT-SUM and MIN is slowed, and COUNT-FROM is skipped entirely. The lack of posterior density over COUNT-FROM actually matches the findings of Siegler and Jenkins (1989), who found that only 1 out of 8 children consistently made use of COUNT-FROM. It is interesting to note, then, that children in Siegler's study were explicitly rewarded for accuracy, but not for speed, suiting the Accuracy Only model.

The reason that speed and accuracy alone can recover the strategy trajectory is because they are both reflective of a common performance pressure: the number of increments required to solve the addition problem. More increments means more opportunity for error (since one increment is one opportunity for error), *and* more increments mean a longer response

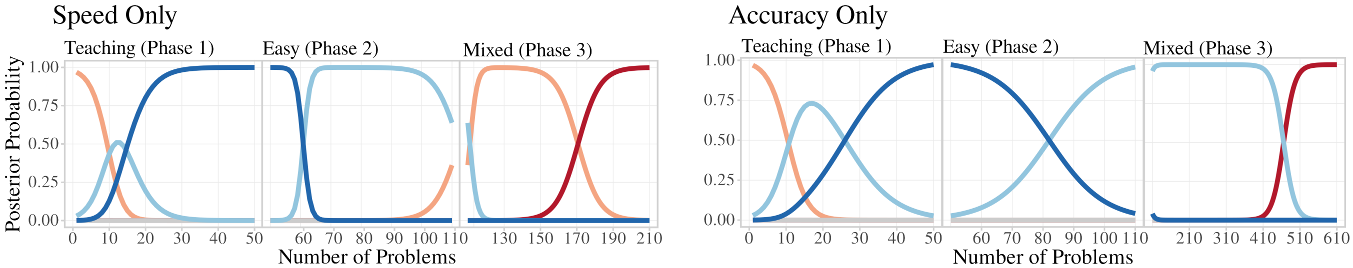

Figure 3: **Speed and accuracy independently produce the trajectory**. Simulations with the three phases of data, but with learners who care only about minimising response time (left) or maximising accuracy (right). In these plots, the qualitative ordering of strategy emergence is similar to each other, the empirical trajectory, and the combined model, but the time taken differs (note the scaling differences on the *x*-axis).

time (since one increment is one unit of additional response time). In both the Speed Only and the Error Only likelihoods, therefore, the strategies are evaluated on qualitatively similar grounds, albeit slightly distinct formalisations.

Discussion

We have demonstrated that the well-known trajectory of strategy discovery in arithmetic learning can be explained by rational inference given a realistic learning environment. Prior models conceptualise the SUM strategy merely as a starting point from which to begin the discovery process, which occurs through internal metacognitive procedures (Jones & Van Lehn, 1994b; Neches, 1987; Shrager & Siegler, 1998). By contrast, a key to our model is that we assume that the process of reconstructing SUM through imitation at the onset of learning plays a central role in children's initial understanding of arithmetic strategy. Consequently, children not only assess whether a strategy fits with performance pressures, but also whether it matches their experience during reconstructive imitation (which favoured SUM). Incorporating these imitation processes results in a cascade through independent learning, producting the four-stage trajectory in the same order that children do given approximately the same number of addition problems (Siegler & Jenkins, 1989).

Until now, we have focused on strategy *discovery*. However, discovery alone is just one aspect of strategy learning. In fact, children often discover a new strategy but do not generalise it to new addition problems until much later (Baroody, 1987; Carpenter & Moser, 1984; Geary et al., 2004; Siegler & Jenkins, 1989). This is the problem of strategy *selection*: for a given problem, children must select which strategy to use from amongst their repertoire. Later in development, children's strategy selection is broadly adaptive (Siegler, 1987): children use strategies like SHORTCUT-SUM on small-numeral addition problems where the likelihood of error is low, and MIN on problems with large and distant numerals, where the advantage of MIN is greatest. Consequently, rational models of strategy selection for addition have achieved some success (Lieder & Griffiths, 2017). A domain for future work is combining our inductive discovery process with these strategy selection selection models (overcoming the limitations of SCADS; Shrager & Siegler, 1998).

Modelling the process of strategy-selection could also explain a dissonance between our induction model and the findings of Siegler and Jenkins (1989). In all but the Accuracy Only simulation, COUNT-FROM is a consistent part of the strategy trajectory in our model; yet in Siegler and Jenkins (1989) just one of eight children used the strategy systematically. What could be occurring is that MIN is not discovered as an entirely distinct algorithm; rather, children use a discovered COUNT-FROM algorithm in conjunction with a strategy selection rule of always starting from the larger numeral. This would require that children know the START-FROM-LARGER selection mechanism before discovering COUNT-FROM. A line of research by Baroody with a much more fine-grained examination of strategy use (reviewed in Baroody & Tiilikainen, 2003) provides evidence that this is the case; the majority of children discover and use START-FROM-LARGER while still at the SHORTCUT-SUM stage, well before discovering COUNT-FROM or MIN. If children discover COUNT-FROM, and use it with their known START-FROM-LARGER rule, MIN shows up empirically. Finally, we note that with the START-FROM-LARGER strategy selection mechanism, searching the space of candidate hypotheses becomes trivial: the only primitive that needs to be in the grammar is count(), and children only need to figure out its arguments. By implementing the context of strategy selection, then, there is even less use for metacognitive rewriting in discovering MIN.

Our work is an important departure from previous accounts of strategy discovery that have used addition as a test case (Crowley et al., 1997a; Jones & Van Lehn, 1994a; Neches, 1987; Rule et al., 2020; Shrager & Siegler, 1998). Mimicking the bias to explain phenomena in terms of the inherent properties of the scientific object (Horne et al., 2025), all previous accounts highlight the learner and their intrinsic metacognitive processes. By incorporating external constraints that shape the environmental distribution an inductive learner observes, we provide an illustration that realistic environmental constraints in combination with general inductive machinery can provide a counter-intuitive explanation of behaviour. One moral of this work, then, is to first look for explanations of behaviour outside of the head before appealing to internal, domain-specific machinery.

References

Almoammer, A., Sullivan, J., Donlan, C., Marušič, F., Žaucer, R., O'Donnell, T., & Barner, D. (2013). Grammatical morphology as a source of early number word meanings. *Proceedings of the National Academy of Sciences*, *110*(46), 18448–18453.

Ashcraft, M. H. (1982). The development of mental arithmetic: A chronometric approach. *Developmental Review*, *2*(3), 213–236. https://doi.org/10.1016/0273-2297(82)90012-0

Ashcraft, M. H. (1992). Cognitive arithmetic: A review of data and theory. *Cognition*, *44*(1), 75–106. https://doi.org/10.1016/0010-0277(92)90051-I

Baroody, A. J., & Tiilikainen, S. H. (2003). Two perspectives on addition development. In A. J. Baroody & A. Dowker (Eds.), *The development of arithmetic concepts and skills constructing adaptive expertise*. Lawrence Erlbaum Associates.

Baroody, A. J. (1984). The Case of Felicia: A Young Child's Strategies for Reducing Memory Demands during Mental Addition. *Cognition and Instruction*, *1*(1), 109–116.

Baroody, A. J. (1987). The development of counting strategies for single-digit addition. *Journal for Research in Mathematics Education*, *18*(2), 141–157.

Carpenter, T. P., & Moser, J. M. (1984). The Acquisition of Addition and Subtraction Concepts in Grades One through Three. *Journal for Research in Mathematics Education*, *15*(3), 179–202. https://doi.org/10.2307/748348

Cowan, R., & Renton, M. (1996). Do They Know What They Are Doing? Children's Use of Economical Addition Strategies and Knowledge of Commutativity. *Educational Psychology*, *16*(4), 407–420. https://doi.org/10.1080/0144341960160405

Crowley, K., Shrager, J., & Siegler, R. S. (1997a). Strategy Discovery as a Competitive Negotiation between Metacognitive and Associative Mechanisms. *Developmental Review*, *17*(4), 462–489. https://doi.org/10.1006/drev.1997.0442

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

Feldman, J. (2000). Minimization of boolean complexity in human concept learning. *Nature*, *407*(6804), 630–633.

Fuson, K. C. (2012). *Children's counting and concepts of number*. Springer Science & Business Media.

Geary, D. C., Hoard, M. K., Byrd-Craven, J., & Catherine DeSoto, M. (2004). Strategy choices in simple and complex addition: Contributions of working memory and counting knowledge for children with mathematical disability. *Journal of Experimental Child Psychology*, *88*(2), 121–151. https://doi.org/10.1016/j.jecp.2004.03.002

Goodman, N. D., Tenenbaum, J. B., Feldman, J., & Griffiths, T. L. (2008). A rational analysis of rule-based concept learning. *Cognitive science*, *32*(1), 108–154.

Gopnik, A., Meltzoff, A. N., & Kuhl, P. K. (1999). *The scientist in the crib: Minds, brains, and how children learn.* William Morrow & Co.

Groen, G., & Parkman, J. (1972). *Psychological Review*, *79*(4), 329.

Groen, G., & Resnick, L. B. (1977). Can preschool children invent addition algorithms? *Journal of Educational Psychology*, *69*(6), 645–652. https://doi.org/10.1037/0022-0663.69.6.645

Heyes, C. (2018). *Cognitive gadgets*. Harvard University Press.

Horne, Z., Kobaş, M., & Cimpian, A. (2025). Historical and experimental evidence that inherent properties are overweighted in early scientific explanation. *Proceedings of the National Academy of Sciences*, *122*(38), e2424725122.

Houlihan, D. M., & Ginsburg, H. P. (1981). The Addition Methods of First- and Second-Grade Children. *Journal for Research in Mathematics Education*, *12*(2), 95–106. https://doi.org/10.2307/748705

Jones, R. M., & Van Lehn, K. (1994a). Acquisition of Children's Addition Strategies: A Model of Impasse-Free, Knowledge-Level Learning. *Machine Learning*, *16*(1), 11–36. https://doi.org/10.1023/A:1022626701243

Jones, R. M., & Van Lehn, K. (1994b). Acquisition of children's addition strategies: A model of impasse-free, knowledge-level learning. *Machine Learning*, *16*, 11–36.

Kaye, D. B., Post, T. A., Hall, V. C., & Dineen, J. T. (1986). Emergence of Information-Retrieval Strategies in Numerical Cognition: A Developmental Study. *Cognition and Instruction*, *3*(2), 127–150. https://doi.org/10.1207/s1532690xci0302_3

LeFevre, J.-A., Sadesky, G. S., & Bisanz, J. (1996). Selection of procedures in mental addition: Reassessing the problem size effect in adults. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(1), 216.

Lieder, F., & Griffiths, T. L. (2017). Strategy selection as rational metareasoning. *Psychological review*, *124*(6), 762.

Lombrozo, T. (2007). Simplicity and probability in causal explanation. *Cognitive Psychology*, *55*(3), 232–257. https://doi.org/10.1016/j.cogpsych.2006.09.006

Marewski, J. N., & Schooler, L. J. (2011). Cognitive niches: An ecological model of strategy selection. *Psychological Review*, *118*(3), 393–437. https://doi.org/10.1037/a0024143

Mollica, F., & Piantadosi, S. T. (2022). Logical word learning: The case of kinship. *Psychonomic Bulletin & Review*, 1–34.

Neches, R. (1987). Learning through incremental refinement of procedures. In D. Klahr, P. Langley, & R. Neches (Eds.), *Production system models of learning and development*. MIT Press.

O'Shaughnessy, D. M., Cruz Cordero, T., Mollica, F., Boni, I., Jara-Ettinger, J., Gibson, E., & Piantadosi, S. T. (2023). Diverse mathematical knowledge among indigenous amazonians. *Proceedings of the National Academy of Sciences*, *120*(35), e2215999120.

Ostad, S. A. (1997). Developmental differences in addition strategies: A comparison of mathematically disabled and mathematically normal children. *British Journal of Educational Psychology*, *67*(3), 345–357. https://doi.org/10.1111/j.2044-8279.1997.tb01249.x

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1988). Adaptive strategy selection in decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(3), 534–552. https://doi.org/10.1037/0278-7393.14.3.534

Piantadosi, S. T. (2014). LOTlib: Learning and Inference in the Language of Thought.

Piantadosi, S. T. (2023). The algorithmic origins of counting. *Child development*, *94*(6), 1472–1490.

Piantadosi, S. T., Tenenbaum, J. B., & Goodman, N. D. (2012). Bootstrapping in a language of thought: A formal model of numerical concept learning. *Cognition*, *123*(2), 199–217.

Rule, J. S., Tenenbaum, J. B., & Piantadosi, S. T. (2020). The child as hacker. *Trends in cognitive sciences*, *24*(11), 900–915.

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. (1987). The perils of averaging data over strategies: An example from children's addition. *Journal of experimental psychology: General*, *116*(3), 250.

Siegler, R. S. (1996). *Emerging minds: The process of change in children's thinking*. Oxford University Press.

Siegler, R. S., & Jenkins, E. (1989). *How children discover new strategies*. Psychology Press.

Svenson, O., & Broquist, S. (1975). Strategies for solving simple addition problems: A comparison of normal and subnormal children. *Scandinavian Journal of Psychology*, *16*(1), 143–148. https://doi.org/10.1111/j.1467-9450.1975.tb00175.x

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. https://doi.org/10.1111/j.1467-9450.1983.tb00483.x

Ullman, T. D., & Tenenbaum, J. B. (2020). Bayesian models of conceptual development: Learning as building models of the world. *The Annual Review of Developmental Psychology is*, *2*, 533–58.

Wynn, K. (1990). Children's understanding of counting. *Cognition*, *36*(2), 155–193.

Xu, F. (2019). Towards a rational constructivist theory of cognitive development. *Psychological Review*, *126*(6), 841–864. https://doi.org/10.1037/rev0000153