One is a dot, two is a pair, three is a trend: a computational model of Quinian bootstrapping

Ria A. Das[1], Samuel J. Cheyette[2], Lionel Wong[3] & Joshua B. Tenenbaum[2]
[1]Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology
[2]Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology
[3]Department of Psychology, Stanford University

Abstract

Conceptual development, both in childhood and across the history of science, is marked by moments of discontinuity: landmark events after which previously inexpressible thoughts and theories, *incommensurable* with the initial cognitive or scientific repertoire, become accessible. Despite longstanding interest, a computational account of these transitions that answers the deepest Fodorian nativist objections remains elusive. Here we present such an explanation by revisiting Carey's *Quinian bootstrapping*. Building upon models based on program induction, we develop an enriched paradigm in which discontinuous conceptual change becomes the invention of a *new type system*, related by *structural analogy* to the previous representation system, but also, crucially, generalizing beyond it. We instantiate our method in the classic domain of children's early mathematical development, modeling the acquisition of both *natural* and *rational number*. We find our model captures several previously unmodeled phenomena, and provides more parsimonious explanations for cross-linguistic universals and intervention effects in number learning.

Keywords: cognitive development; language of thought; bootstrapping; number word learning; concepts; program induction

Introduction

A longtime dream of research in computational cognitive science is achieving a formal model of children's cognitive development: How do children grow increasingly rich systems for representing and reasoning about the world as they age? This evolution is shaped by two knowledge systems developed over longer timescales. The first is *biologically innate* representation systems, such as functionally-specific brain regions that implement systems of *core knowledge*, developed over evolutionary time and common across cultures. The second is *cultural* knowledge systems, such as symbolic language and mathematics, developed by human civilizations over the course of historical time. Here, cross-cultural differences can lead to variation in the pace and outcomes of children's developmental trajectories, sometimes leading to adult populations with markedly different cognitive abilities (Frank et al., 2008).

In this paper, we focus on a particularly transformative kind of developmental change known as a *discontinuous conceptual* change. These transitions are landmark events after which previously inexpressible thoughts and theories, *incommensurable* with the initial cognitive repertoire, become accessible. An archetypal example is children's acquisition of number concepts, where the developmental signature across industrialized cultures is children's sudden understanding of all number words in their count list after initial incremental learning. Remarkably, evidence from cognitive-historical analysis suggests that the cognitive mechanisms underlying these childhood discontinuities share many similarities with those metacognitively used by scientists who have created discontinuity in *scientific knowledge*, like Kepler and Darwin (Carey, 2009). A unified model that crystallizes the mechanism operating in both of these settings would be a significant achievement.

Despite longstanding interest, such a satisfying computational explanation remains elusive. Past attempts to model conceptual change based on *program induction* (e.g., Piantadosi et al., 2012, Ellis et al., 2021) have made key progress, explaining how new structured representations can be learned. However, they have not addressed the deepest nativist objections to discontinuity, because they build in initial access to the mental primitives—states and functions—needed to express the conceptual system ultimately developed (Fodor, 1998).

Here we present an algorithmic-level explanation of how genuinely new conceptual systems can be constructed via a formalization of *Quinian bootstrapping*, the informal class of mechanisms posited by Carey to explain radical conceptual change (Carey, 2009). Taking inspiration from a modern approach to the language-of-thought (LoT) hypothesis, we model a child's mental representation system for a given domain at a given time as a formal *programming language*—a *domain-specific language* or *DSL*—and conceptual change over time as *DSL synthesis*. Unlike previous LoT-based models, we

1. enrich the semantic search space beyond just functions over *fixed* primitive states to include *new* representations of state or *types*, and
2. explicitly model *analogical mapping* as part of the learning signal via the discovery of shared *relational structure* between the current hypothesis and some external inputs, such as symbolic language, or representational substrate.

In this richer paradigm, discontinuous ("Quinian") conceptual change becomes the invention of a *new type system*, defined over the external symbols and related by analogy to the previous system, but also, crucially, generalizing beyond it.

We evaluate our approach in the classic domain of children's early mathematical development and model the acquisition of both *natural number* and *rational number* as novel conceptual types discontinuous with previously available forms. As referenced, after learning the meanings of number words up to three or four, children famously undergo a sudden and consistent leap to understanding all words in their count list around age 3.5 (the cardinal principle or CP transition). Later, they enrich this understanding after learning fractions (typically between ages 8 and 12). We show that our account reproduces Carey's proposed mechanisms for each stage of this learning trajectory, capturing several previously unmodeled phenomena. In the natural number domain, these newly modeled observations include (1) the multiple stages of the often atomically-viewed CP induction, and (2) cross-linguistic diversity in learning trajectories caused by morphological differences in quantifier structure. In the rational number setting, they include (1) intrusions from the natural number domain in early stages of arithmetic learning and (2) curriculum effects on the rate of fraction learning. In addition, our model provides more parsimonious explanations than existing methods for cross-linguistic universals, such as the steady three-number prerequisite for the CP induction, and intervention effects.

More generally, our work provides a computational-level grounding of a longstanding theory on the human capacity for genuine conceptual change—the ability to think truly new thoughts—as a cascading of links from old to new type systems guided by generative analogies. In the rest of the paper, we present a detailed example and related work; formalize the model; and discuss results in the two number domains.

Example and related work

In plainest terms, Quinian bootstrapping is a mechanism that generalizes an initial conceptual system (CS1) into a more expressive conceptual system (CS2), by *adding meaning to the relations* in a *placeholder structure*: externally-provided symbolic structure such as natural language (Carey, 2009). Concretely, we model each conceptual system as a set of interrelated *type systems*, which may be viewed as miniature DSLs that each define the semantics of a single *type*. A type is defined by a set of *states* that may be used to represent the world, as well as a set of *operations* that describe valid relations between states. Further, CS1 additionally includes a representation of placeholder structure as an *incomplete* type system, or a set of external symbols with undefined or partially defined meanings for both its states and operations.

In the natural number setting (Figure 1), CS1 is composed of two type systems, corresponding to the two core number systems: the parallel individuation (PI) system and the approximate number system (ANS). The placeholder structure is the *verbal counting sequence*, i.e. the placeholder symbols are the number words, and the key placeholder relation is the *next_word* relation. Prior to learning the meanings of number words, children have already memorized the counting sequence up to a highest number, but critically, this knowledge is limited to knowing that each number comes *after* the previous one, not that each is *one more*. Enriching this limited understanding of the *next* relation, or put differently, *adding meaning* to it, is key to reaching the CP induction, modeled as transforming the incomplete type system representing the placeholder structure into a complete one with fully defined semantics and interrelations with the other type systems.

How does this work? First, the meanings of the first three (sometimes four) number words are learned in terms of the parallel individuation system, driven by the utility of knowing these meanings for performing tasks in the world (details provided in the following section). At this stage, however, the *analogy* between the *next_word* relation in the symbol (word) domain and the *add_one* relation in the physical (meaning) domain has not yet been discovered.

Unlike the earlier one-knower and two-knower stages, however, the three-knower stage is uniquely ripe as a substrate for making this discovery. This is because there is a *suspicious coincidence*: two *next_word* relations in the symbol domain align with two *add_one* relations in the physical domain. To use the language of the field of programming languages, noticing such a suspicious coincidence reveals an opportunity for *compression* of the mental representation system. Namely, rather than memorizing the count sequence and *separately* representing the meanings of the first three number words as three totally *unrelated* PI states, the full conceptual system can be more efficiently represented as (1) the count sequence, (2) the meaning of one, and (3) meanings of two and three *equivalently* defined as evaluations of *add_one* performed *in lockstep* with the already remembered count list.

Naturally, following analogy discovery, the next and final step of Quinian bootstrapping is analogy *generalization*. Unifying the relationship between the meanings of (one, two) and (two, three) as described suggests that a similar semantic relationship may exist between the meanings of later consecutive pairs in the count list. However, naively extending the analogy in this way fails after the meaning of four, since this is the upper limit of parallel individuation, and the *add_one* relation previously used is an operation supported and thus limited by that system. Put differently, the ability to represent exact quantities above four does not exist in either of the two core number systems, so the meanings of all the counting words cannot be represented by using only these systems.

This dilemma brings us to the deepest contribution of Quinian bootstrapping: Despite the limitations of the domain-specific core knowledge systems, truly novel conceptual invention can still occur via the deployment of *domain-general* cognitive constructs—like the ability to define *recursive* data types or representations of state—that *generalize* the relations supported by core cognition beyond their limits. While these constructs are inherently more difficult to access than core knowledge, partly due to their infinite variety and children's lack of practice with them, evidence suggests that they become more accessible when suggested by the *symbolic structure of the environment*, be it physical or *abstract*—e.g., in the form of the current mental representation system and its model of the externally-provided placeholder structure.

Precisely, the *recursive* structure of the analogy between the count list and the meanings of the first three number words suggests the invention of a new *recursive type* using the placeholder symbols. In particular, rather than attempt-

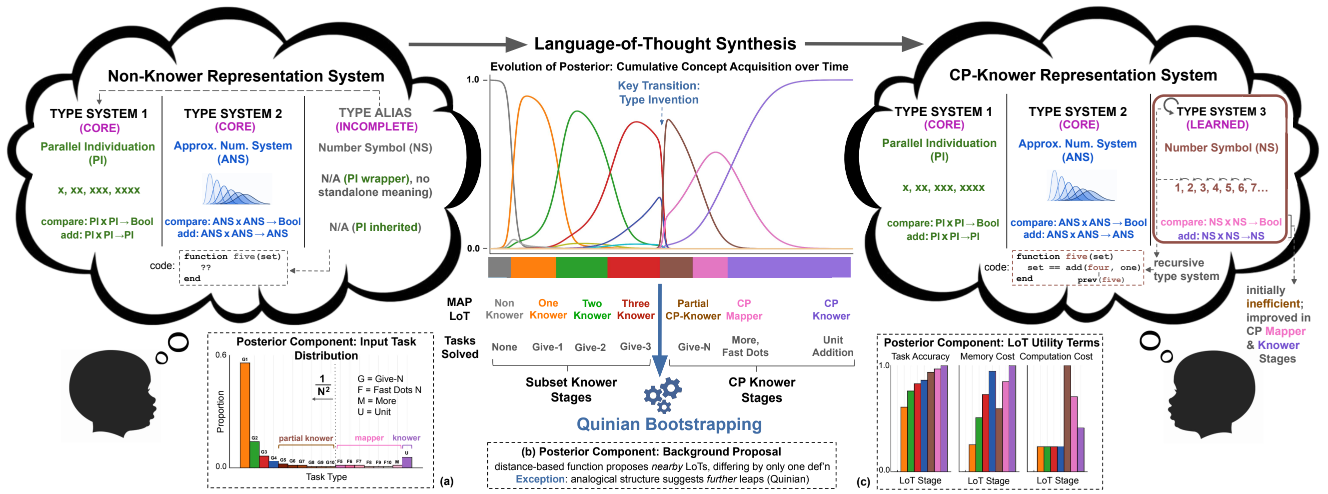

Figure 1: Overview of Natural Number Model. (a), (b), and (c) are all inputs to inference (unrelated to side of diagram).

ing to *evaluate* the PI-based *add_one* operation to create an *iconic* mental representation of sets greater than size four (i.e. beyond normal working memory capacity), the mental representation of a number word like five may just be represented as "one more than four," or more generally, "one more than the previous word's meaning." Concretely, this is modeled by the invention of a new operation in the placeholder domain,

$$add : NS \times NS \rightarrow NS,$$

that generalizes the analogous relationship in the PI type system to all known number symbols (*NS*), but without requiring the imagination of *any* iconic or otherwise physically-grounded mental format. Instead, it is defined using just the lightweight *next_word* function. Further, the placeholder type definition is updated so that the symbols are no longer mere *wrappers* of other type systems for representing quantities, but a novel, recursively-defined type system that itself captures unique quantity representations—namely, quantities defined using *relation expressions* like *add(four, one)*. Critically, this new symbolic type system is fully *detached* from the physical realm, enabling reasoning about exact discrete quantities without explicitly imagining sets of different sizes (though it can always be *translated back* to physical sets by explicitly evaluating the recursion). As such, it is the key stepping stone towards efficiently reasoning about higher-order symbolic mathematics in later education.

Formally, we implement this two-step algorithm of (1) analogy discovery and (2) analogy generalization as a search for *relational compressibility* in the knowledge representation. We assume that, over the course of their lives, children and adults sometimes notice common relational patterns between parts of their knowledge representation and external signals. This recognition can then lead to knowledge *compression* and *generalization*. Unlike standard compression algorithms that search for common structure across the *semantics of individual functions* ("functional compression"), our model searches for common structure across the *relations between semantics*, be they functions or states, capturing how *grounded* a knowledge system is by an underlying network of analogies ("relational compression"). Overall, we wrap this analogy search method in a utility-based rational learning paradigm (described next), where relational compressibility decreases the *memory cost* of storing a knowledge representation and the *discovery cost* of proposing a new one from the current via generalization.

Main related work: Piantadosi et. al. 2012, 2023

To conclude this overview, we briefly describe the differences between our model and the closest previous work, the rational statistical inference (RSI) model of number learning by Piantadosi, 2023. Our model makes three changes with respect to RSI that let us better explain empirical data, and resolve Carey's critiques about why it is not Quinian bootstrapping.

First, we model each learning stage as being *dependent* on the previous one, in that the *structure* of the current LoT may influence search for new LoTs. In contrast, RSI models learning stages as being fully *independent*, so that the timing of the CP induction is driven by the low prior probability assigned to the *recurse* primitive in the search space. As such, our model provides a *structural* reason why the CP induction always occurs after at least the three-knower stage, and never, say, the two-knower stage. Namely, the relational shared structure between the count list and children's initial PI-based definitions of the first three numbers is a *suspicious coincidence* that draws attention to the key analogy, triggering compression and generalization to the full count list. In contrast, lightly tweaking the RSI prior assigned to *recurse* could induce the CP stage after the second or earlier stages, a pattern not observed, even in cultures where children experience the CP induction at much older ages (when the general ability to note recursion would be expected to have improved, modeled as a potentially higher prior given to *recurse*; Piantadosi et al., 2014).

Second, the external learning signal in the RSI model comes exclusively from the distribution of individual number words.

As such, it predicts that adding more low-number (1-3) counting input at the three-knower stage will not accelerate the CP induction, because it does not affect the LoT's accuracy (three-knowers already know all those word meanings). Counting intervention experiments, however, show this is false: added low number counting *does* accelerate the CP induction (Gibson et al., 2020). Our model predicts this, because more low-number counting input at the three-knower stage accelerates search via *further analysis* of the current hypothesis, causing faster discovery of the analogy leading to the CP leap.

Third, rather than modeling the evolution of a *single function*, namely, the counting algorithm, we model a child's domain representation as a *related set of functions and states*. By carefully modeling the initial primitive states and functions of core numerical cognition, and distilling domain-specific and domain-general biases, this richer representational substrate gives us a platform to *jointly* model how children learn all the *concepts* and *algorithms* of higher math. We instantiate the first step of this research program in this paper by modeling how children augment their hard-won understanding of natural number with the significantly richer notion of *rational number*. There, evidence from educational studies suggests that deeper understanding is also driven by the discovery of a key *symbolic-physical analogy*, as formalized by our approach.

Model

We now present a formal model of our learning system. We frame conceptual evolution as *search* for a mental representation system (LoT) L within a space $\mathcal{L}$, guided by (1) the *utility* of L for solving a distribution of reasoning tasks T (Figure 1a), and (2) the difficulty of *discovering* L from an initial mental representation system L_0. In detail, the utility $U(L;T)$ balances a task *accuracy* term, $A(L;T)$, with two *cost* terms, a *memory cost* $C_m(L)$ capturing how difficult L is to remember (based on description length and relational compressibility), and a *computation cost* $C_c(L;T)$ capturing how *efficiently* L solves the tasks T (based on the length of its *execution trace*):

$$U(L;T) = \gamma_1(t)A(L;T) - \gamma_2 C_m(L) - \gamma_3 C_c(L;T) - \gamma_4.$$

The time-dependent parameter $\gamma_1(t)$ controls the tradeoff between cost and accuracy, increasing over time to model growing tolerance for higher-cost representations if they are also highly accurate. Further, the *discovery cost*, $C_d(L;L_0,T,t)$, is used as part of a *background proposal function* (Figure 1b) that proposes new LoTs L from the current state L_0:

$$P(L \mid L_0;T,t) \propto \exp(-\alpha C_d(L;L_0,T,t)).$$

The added dependence on the task distribution T and time t reflects an increasing ability to propose more *distant* (e.g. CP-knower is more distant from three-knower than four-knower is) LoTs L from the starting state L_0 over time (t), which may be further accelerated by increasing exposure to particular kinds of inputs in the task distribution (T). This design explains the accelerated pace of learning caused by low-number counting interventions—a modification of the base task distribution T—and *curriculum* effects in the rational number domain.

These utility and discovery components are unified in a single learning paradigm using the formalism of rejection sampling. Namely, the discovery-cost-based background proposer is used to *suggest* new possible LoTs, and the utility function is used to decide whether to *accept* or *reject* the proposal:

$$a(L;T,U^*) = \min\{1, \exp(\beta[U(L;T) - U^*])\}, \text{ with}$$

$$U^* \geq \max_L U(L;T).$$

Finally, the posterior is proportional to the product of the background proposal and the utility-based acceptance probability.

Overall, this model is different in three ways from more standard formulations of utility-based models. (1) While standard formulations balance a single *complexity* cost term against an accuracy term, we split the complexity cost into *two* terms: a discovery cost capturing the difficulty of *proposing* a hypothesis, and a memory cost capturing the difficulty of *remembering* a hypothesis once proposed. This distinction captures the intuition that, while the CP induction is hard to *discover*, it is not that hard to *remember*, with fewer distinct definitions than the higher-knower stages due to relational understanding. (2) We add a computational cost term, modeling the observation that computational efficiency on tasks can also drive conceptual evolution (see results on modeling the multiple stages of the CP induction). (3) Traditional definitions of complexity cost are based on the description length of a hypothesis, where the final length is often computed after a *functional compression* step that shortens the length by extracting common structure in function semantics. Our model extends this by accounting for *relational compressibility*, or the existence of shared structure in the *relations between semantics*, capturing the notion that knowledge systems grounded by common relations are often easier to store (memory cost; Figure 1c) and make particular generalizations easier to find (discovery cost; Figure 1b).

Results

We now describe the results of instantiating our model.

Domain 1: Natural Number

In our current natural number implementation, the LoT search space $\mathcal{L}_N$ contains 75 LoTs, including the key observed stages (i.e. one-knower, two-knower, etc.), along with other definitions that children might assign to the number words one through ten based on parallel individuation (e.g. two means x), the ANS (e.g. one means *approximately one*), or a combination of both (e.g. one means x, and each subsequent number is approximately more than the last). A subset of these LoTs are listed in the center-right legend of Figure 2.

Inducing the Cardinal Principle The base result—the evolution of the incremental subset-knower stages leading to the CP induction—is shown in Figure 1. As the cost tolerance increases over time, higher-cost but also higher-accuracy LoTs emerge. The key Quinian transition takes place between the three-knower and *partial CP knower* stage once the discovery

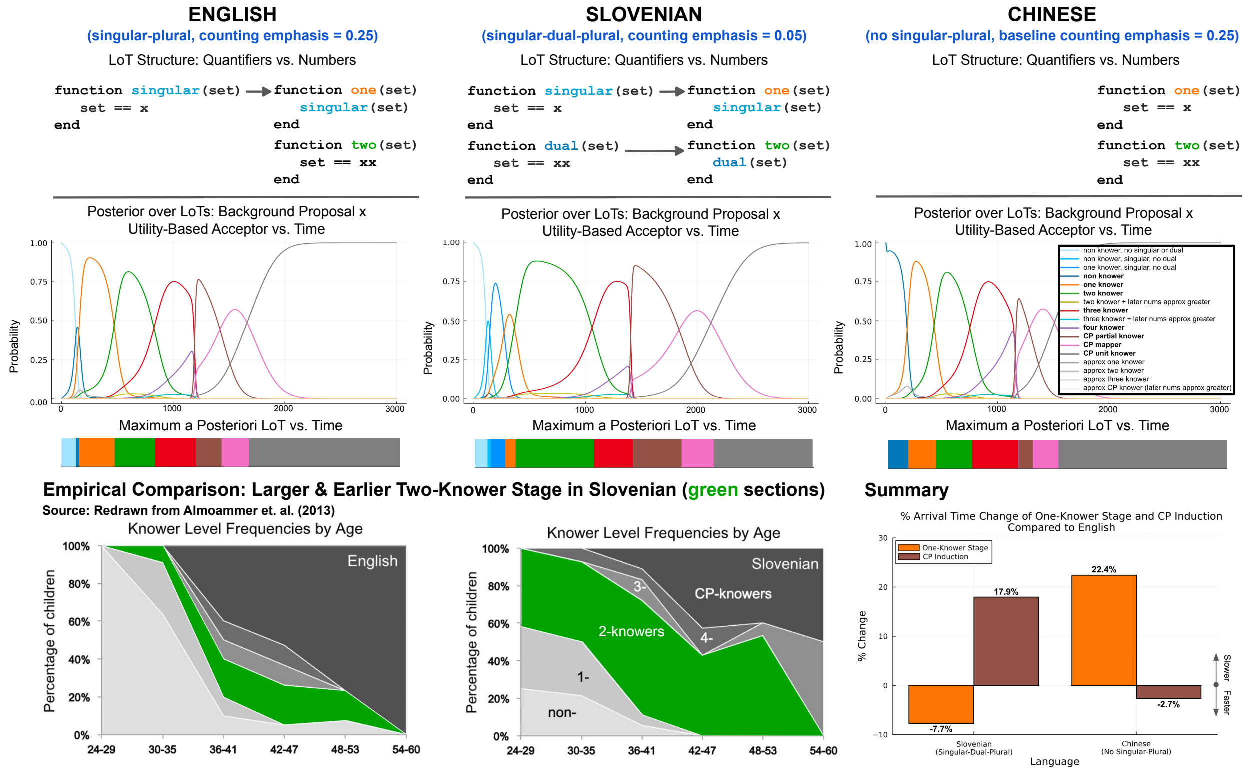

Figure 2: Cross-Linguistic Diversity Results. The direction, not magnitude, of change is most relevant in this prototype model.

cost, modeled as *decreasing* in proportion to the time t as children *increasingly* analyze the three-knower LoT's structure, is low enough for the CP leap to be realistically proposed.

Multiple CP Stages: Partial Knower, Mapper, and Knower While traditionally viewed as an atomic event, Davidson et al., 2012 and others have shown that many children who solve the Give-N task, historically the CP-knower criterion, *cannot* answer other basic number questions. These other tasks include the More task, which asks which of a pair of higher numbers is larger ("Which is more, six or nine?"); the Fast Dots task, which asks children to estimate the number of dots in a flashed set, with the correctness criterion being stating an increasing list of numbers for increasingly large sets (indicating knowledge of the *later-greater* principle); and the Unit task, which tests children's understanding of the exact unit difference between consecutive number words ("If I add one to seven, do I get eight or nine?"). Previous models have failed to capture these nuances, because they only model the acquisition of one part of number understanding, a single function representing the counting algorithm, as opposed to the full conceptual system also describing relative magnitude understanding (More and Fast Dots tasks) and exact arithmetic (Unit task).

Our richer modeling formalism lets us explain this behavioral data by representing the post-Give-N concepts as *additional function semantics* to learn upon recognizing natural numbers as a recursive type system. In the language of computer science, these extra functions—namely, the *compare* and *add* functions—form a *programming interface* for any type system for quantity representation, meaning that bestowing that label upon number words requires having to define these interface functions as well (indeed, both core number systems include built-in semantics for these functions).

Concretely, following a hypothesis proposed in Davidson et al., 2012, we model the initial, Give-N-based CP stage (partial CP-knower) as including *default* definitions for both functions that are *overly grounded* in the newly-understood counting procedure. Precisely, to *compare* two numbers above the PI limit, children at this stage *count up* from one for each number, and then give an answer. Similarly, to *add one* to a higher number, they also count up from one to the largest number, and then one more—despite being able to answer what the *next* word is immediately. Both of these methods are tremendously *computationally inefficient*, causing children to give up and guess, especially at higher numbers. Conceptual evolution takes place as children learn more efficient algorithms for comparison and addition (similar to Shrager and Siegler, 1998), by memorizing the later-greater principle (CP-mapper), and later refining it via the exact +1 principle (CP-knower). While memorizing these extra facts *increases the memory cost*, the *gain in accuracy* and *reduction in computational cost* (Figure 1c) eventually causes them to emerge.

Cross-Linguistic Diversity A longstanding hypothesis in number learning is that the morphology of natural language might play a role in how children assign initial meanings to

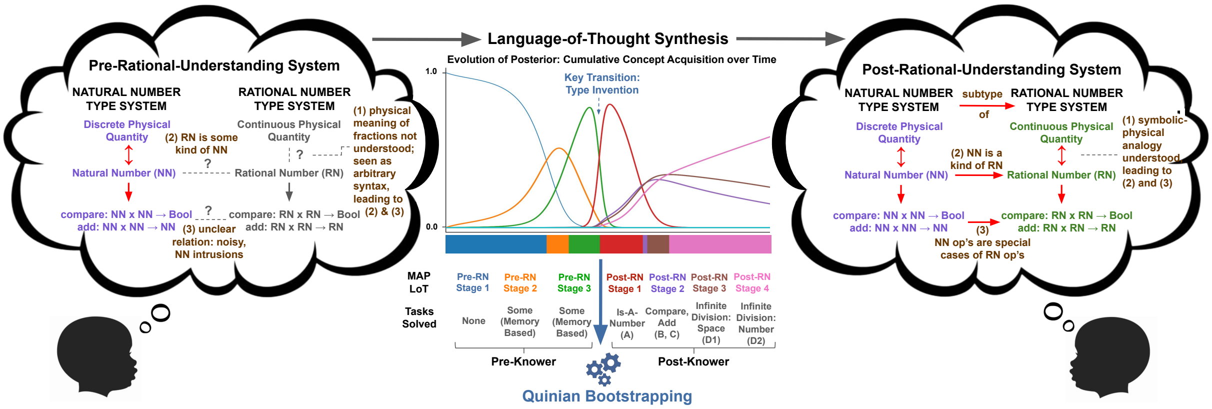

Figure 3: Overview of Rational Number Model.

number words (Bloom and Wynn, 1997). In particular, the *quantifier structure* of language, such as singular-plural syntax, may provide a learning signal, via structures like singular articles and nouns—which are higher frequency in children's language input and thus learned earlier—being used in the same context as number words like "one." This predicts that children raised in languages with different quantifier structures may exhibit different paces of learning. In particular, children taught languages with *richer* quantifier structure might be expected to learn the meanings of the first few number words *sooner*. This prediction has been empirically corroborated with children's number learning data across several languages (e.g. Sarnecka et al., 2007, Villarroel et al., 2011), but has not been explicitly modeled in existing formalisms.

Our model naturally captures these effects, because it supports the addition of functions corresponding to singular, dual, and plural marker assignment to the knowledge representation, as well as a learning mechanism where the existence of definitions for these terms in the current hypothesis increases the probability of *discovering the same definitions* for number words, namely one (singular) and two (dual), as shown in Figure 2. By (1) adding these quantifier representations to the LoT for each natural language, (2) modifying the input task distribution to reflect better practice with early number words when their meanings are emphasized by hints in quantifier structure, and (3) incorporating variant emphases on *counting instruction* across different languages—approximated by the highest count of same-age children raised in different cultures—as a linear coefficient affecting the discovery cost of the CP generalization from the three-knower stage over time, we explain observed differences in number learning trajectories across several languages (Le Corre et al., 2016, Almoammer et al., 2013). We find that children raised in languages *without* singular-plural structure (Chinese) are *slower* to learn the meaning of one compared to those raised in languages *with* singular-plural structure (English), and children raised in languages with singular-*dual*-plural structure (Slovenian) learn the meanings of one and two *faster*, without the same-direction effects on the ultimate arrival of the CP induction.

Domain 2: Rational Number

The key idea formalized by our model, implemented over a space $\mathcal{L}_R$ of 110 LoTs describing variant semantics of fractions and their operations, is that children originally do not see fractions as measures of continuous physical quantities, but instead as *special kinds* of natural numbers (Moss and Case, 1999). This leads to several mistakes, especially caused by natural number intrusions—the existing and more relationally grounded knowledge representation—in performing fraction comparison and arithmetic (e.g. $\frac{1}{2} + \frac{1}{3} = \frac{2}{5}$). This misconception is corrected upon discovering the symbolic-physical analogy between fraction syntax and physical quantities (Smith et al., 2005), which may be accelerated by curricula that emphasize analogy-related tasks versus memorization-based curricula (Carey, 2009). For space reasons, we include just the key result (Figure 3), leaving detailed results to this OSF link.

Discussion

To end, how does our method overcome the original nativist critique? Our key idea is to separate *domain-specific cognition*—initial systems like core knowledge (e.g. PI, ANS) that are immediately accessible, but have limited expressivity—from *domain-general cognition*—an infinite grammar capturing the arbitrary complexity of symbolic thought, including new abstractions of state (e.g. Quinian refactorings of types), but which is difficult to access. Prior approaches manage this dichotomy by mixing domain-specific primitives with a *hand-selected* set of domain-general ones to create a single, problem-specific search space, where the latter primitives can be discovered at *any point* in learning, *independent* of the current hypothesis (Piantadosi et al., 2012). In contrast, we argue that a model in which domain-general constructs *become accessible* when they are *suggested by the symbolic structure of the cultural environment*—especially cultural inventions like the number list or fraction syntax—is a more parsimonious explanation for empirical learning data, and avoids the initial primitive "build-in" noted by Fodor. We look forward to developing this proof-of-concept further in future work.

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. https://doi.org/10.1073/pnas.1313652110

Bloom, P., & Wynn, K. (1997). Linguistic cues in the acquisition of number words. *Journal of Child Language*, *24*(3), 511–533. https://doi.org/10.1017/S0305000997003188

Carey, S. (2009). *The origin of concepts* (1st). Oxford University Press, Inc.

Davidson, K., Eng, K., & Barner, D. (2012). Does learning to count involve a semantic induction? *Cognition*, *123*(1), 162–173. https://doi.org/https://doi.org/10.1016/j.cognition.2011.12.013

Dechter, E., Malmaud, J., Adams, R. P., & Tenenbaum, J. B. (2013). Bootstrap learning via modular concept discovery. In F. Rossi (Ed.), *IJCAI 2013, proceedings of the 23rd international joint conference on artificial intelligence, beijing, china, august 3-9, 2013* (pp. 1302–1309). IJCAI/AAAI. http://www.aaai.org/ocs/index.php/IJCAI/IJCAI13/paper/view/6890

Ellis, K., Wong, C., Nye, M., Sablé-Meyer, M., Morales, L., Hewitt, L., Cary, L., Solar-Lezama, A., & Tenenbaum, J. B. (2021). Dreamcoder: Bootstrapping inductive program synthesis with wake-sleep library learning. *Proceedings of the 42nd ACM SIGPLAN International Conference on Programming Language Design and Implementation*, 835–850. https://doi.org/10.1145/3453483.3454080

Fodor, J. (1998, February). Bibliography. In *Concepts: Where cognitive science went wrong*. Oxford University Press. https://doi.org/10.1093/0198236360.001.0001

Frank, M. C., Everett, D. L., Fedorenko, E., & Gibson, E. (2008). Number as a cognitive technology: Evidence from pirahã language and cognition. *Cognition*, *108*(3), 819–824. https://doi.org/https://doi.org/10.1016/j.cognition.2008.04.007

Fuson, K. C. (1998). *Children's counting and concepts of number*. Springer Verlag.

Gelman, R., & Gallistel, C. (1978). *The child's understanding of number*. Harvard Univ. P.

Gibson, D. J., Gunderson, E. A., & Levine, S. C. (2020). Causal effects of parent number talk on preschoolers' number knowledge. *Child Development*, *91*(6), e1162–e1177. https://doi.org/10.1111/cdev.13423

Le Corre, M., Li, P., Huang, B. H., Jia, G., & Carey, S. (2016). Numerical morphology supports early number word learning: Evidence from a comparison of young mandarin and english learners. *Cognitive Psychology*, *88*, 162–186. https://doi.org/https://doi.org/10.1016/j.cogpsych.2016.06.003

Moss, J., & Case, R. (1999). Developing children's understanding of the rational numbers: A new model and an experimental curriculum. *Journal for Research in Mathematics Education*, *30*(2), 122–147. Retrieved January 31, 2026, from http://www.jstor.org/stable/749607

Piantadosi, S. T. (2023). The algorithmic origins of counting. *Child Development*, *94*(6), 1472–1490. https://doi.org/https://doi.org/10.1111/cdev.14031

Piantadosi, S. T., Jara-Ettinger, J., & Gibson, E. (2014). Children's learning of number words in an indigenous farming-foraging group. *Developmental Science*, *17*(4), 553–563. https://doi.org/https://doi.org/10.1111/desc.12078

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. https://doi.org/https://doi.org/10.1016/j.cognition.2011.11.005

Sarnecka, B. W., Kamenskaya, V. G., Yamana, Y., Ogura, T., & Yudovina, Y. B. (2007). From grammatical number to exact numbers: Early meanings of 'one', 'two', and 'three' in english, russian, and japanese. *Cognitive Psychology*, *55*(2), 136–168. https://doi.org/https://doi.org/10.1016/j.cogpsych.2006.09.001

Shrager, J., & Siegler, R. S. (1998). Scads: A model of children's strategy choices and strategy discoveries. *Psychological Science*, *9*(5), 405–410. https://doi.org/10.1111/1467-9280.00076

Smith, C. L., Solomon, G. E., & Carey, S. (2005). Never getting to zero: Elementary school students' understanding of the infinite divisibility of number and matter. *Cognitive Psychology*, *51*(2), 101–140. https://doi.org/https://doi.org/10.1016/j.cogpsych.2005.03.001

Villarroel, J. D., Miñón, M., & Nuño, T. (2011). The origin of counting: A study of the early meaning of 'one', 'two' and 'three' among basque- and spanish-speaking children. *Educational Studies in Mathematics*, *76*, 345–361. https://doi.org/https://doi.org/10.1007/s10649-010-9291-0