RESEARCH ARTICLE

Compositional Reasoning in Early Childhood

Steven Piantadosi*, Richard Aslin

Department of Brain and Cognitive Sciences, University of Rochester, Rochester, NY, United States of America

* spiantado@gmail.com

OPEN ACCESS

Citation: Piantadosi S, Aslin R (2016) Compositional Reasoning in Early Childhood. PLoS ONE 11(9): e0147734. doi:10.1371/journal.pone.0147734

Editor: Amanda Bruce, University of Kansas Medical Center, UNITED STATES

Received: November 23, 2014

Accepted: January 7, 2016

Published: September 2, 2016

Copyright: © 2016 Piantadosi, Aslin. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are available via Figshare (<https://dx.doi.org/10.6084/m9.figshare.3492779.v1>).

Funding: Research reported in this publication was supported by the Eunice Kennedy Shriver National Institute of Child Health & Human Development of the National Institutes of Health under Award Number F32HD070544 and an NIH research grant to R. Aslin and E. Newport (HD-037082). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Competing Interests: The authors have declared that no competing interests exist.

Abstract

Compositional “language of thought” models have recently been proposed to account for a wide range of children’s conceptual and linguistic learning. The present work aims to evaluate one of the most basic assumptions of these models: children should have an ability to represent and compose *functions*. We show that 3.5–4.5 year olds are able to predictively compose two novel functions at significantly above chance levels, even without any explicit training or feedback on the composition itself. We take this as evidence that children at this age possess some capacity for compositionality, consistent with models that make this ability explicit, and providing an empirical challenge to those that do not.

1 Introduction

One of the basic goals of cognitive science is to understand the nature of conceptual representations. Proposals for the format and structure of concepts range through theories based on prototypes [1], exemplars [2, 3], language-like representations [4], and others [5]. A central issue for all approaches is that of *compositionality*—how to handle the combination of individual concepts into more complex representations (e.g. “beautiful” composed with “accordions” yields a composite concept, “beautiful accordions”) [6–9]. This ability may be built “on top” of a representational format like prototypes [10] or it may be intrinsically a part of the core conceptual representation itself, as is hypothesized in *language of thought* (LOT) theories. According to LOT theories, the formation of a complex thought like “all sailboats are heavy” occurs when thinkers construct the corresponding compositional structure built out of simpler operations. For instance, such a conceptual format may be similar to standard logic: $\forall x.sailboat(x) \rightarrow heavy(x)$. Here, $\forall$, *sailboat* and *heavy* are functions that are composed in one particular way to express the composite idea “all sailboats are heavy.” These functions may themselves be built out of simpler operations, but in LOT theories eventually this process grounds out in primitive functions from which all other thoughts and representations are derived.

LOT theories are motivated in part by the rich structure human concepts appear to easily represent, as well as by factors like the systematicity, productivity, and compositionality of cognition [11]. Indeed, the LOT has a rich history in cognitive science, dating back at least to Boole [12] and later Frege [13]. Boole sought to characterize the “laws of thought” and did so by postulating the formal logical system of Boolean logic, which builds logical representations (formulas) out of the logical primitives *and*, *or*, and *not*. More recent incarnations of LOT

theories have argued specifically for compositional, symbolic representations as fuller cognitive theories encompassing a wider range of semantic and logical operations [4]. A structured LOT has also been argued to explain developmental change in concept and language learning [14–21], providing a convenient formalism to express both what learners bring to learning tasks (primitive functions), and what exactly they acquire (compositions of those functions to explain observed data). For instance, Piantadosi et al. [21] suggest that children may know about simple functions on sets (e.g. union, intersection, etc.) and build more complex representations of cardinality and counting by appropriately composing these operations. This type of general approach builds on work of Feldman [22] who used a compositional system to explain patterns in Boolean concept learning, and Siskind [14] who first developed a compositional account of learning semantics. Following Goodman et al. [16], many recent theories posit rational (Bayesian) statistical models as the core *inductive* mechanism that decides between possible combinations of primitives in the face of data. In all cases—dating back to Boole—the core claim is that composition of simple primitive functions is what allows for the creation of complex cognitive representations.

While learning studies and modeling work have established LOT-based models as plausible *in principle*, no experimental work has examined LOT models’ most basic assumptions: learners should be able to represent functions and combine them through composition. This ability should be present without any explicit instruction on either the fact that functions can be composed, or what happens when they are composed. This ability is not trivial. Functions may perform many kinds of operations and in principle function combination can be used to express arbitrary computation [23–25].

To be clear, we focus here on *conceptual* compositionality (combinations of concepts), not *linguistic* compositionality (combinations of words in sentences), which has been studied previously [26–29]. The reason for this is that acquisition models have so far focused on learning either individual word meanings by composing simpler conceptual elements [14] or using compositional LOT systems to learn non-linguistic systems of concepts like magnetism [19]. In these cases, the learning models therefore make the strong prediction that children should be able to compose operations outside of the domain of language. Note that an ability to compose linguistic expressions does not establish that children will be able to compose conceptual operations; the compositionality of linguistics may or may not be encapsulated within language processing. The present work takes a first step towards investigating non-linguistic conceptual combination in children.

Of course, it is important to note that many tasks can be *interpreted* as compositional, depending on how they are formalized. Even, for instance, Sally-Ann [30] or simple numerical tasks with infants [31] require multiple updates to a representation, and therefore can be phrased as compositions of functions. Here, however, we approach compositionality from a more directed stance, constructing an experiment that is difficult to interpret in any way but compositional. The general logic of our experiment is to teach children that certain “screens” (occluders) cause specific feature changes (color or pattern) to an occluded object. The critical test conditions occur when two such screens are adjacent and children observe an object go behind both screens but they do not observe it between screens. Prediction of the correct outcome in this case requires them to apply both object transformations. Children receive no feedback on their responses to the compositional (two-screen) trials, meaning that success is indicative of *automatic* compositional reasoning without instruction or encouragement.

This experiment tests whether children can compute a composition of functions like $f_2(f_1(x))$, where f_2 is the second screen and f_1 is the first. However, there is a stronger sense in which learners might “know” compositionality: they might be able to form an explicit symbol $h = f_2(f_1(x))$ for the composition itself. Creating such a symbol is not a necessary ability for

success on our task. In a linguistic analogy, the strong form of compositionality is tantamount to knowing that the term (symbol) “pit stop” refers to “refueling” and “changing tires.” The weaker form corresponds to being able to refuel and then change tires, without knowing that there is a term or symbol that refers to the combination of both operations. Because we are just beginning to explore the type of compositional reasoning in a non-linguistic task, we start by investigating the developmental origins of the weaker sense: can preschoolers predictively compose novel operations? Or, do they show catastrophic failures when attempting to combine multiple functions, perhaps reminiscent of younger children’s limitations with multiple updates to representations of objects or numerosity [32–34]?

2 Methods

In the experiment, preschoolers were shown displays on a touchscreen monitor in which a car with a colored pattern appeared. Cars had one of two patterns (dots/stars) occurring in one of two colors (red/blue). A car with one choice of these features drove on-screen and children were required to select which pattern it matched from a menu with all four possible options (Fig 1a). This ensured that they attended to the features of the cars and knew them at the start of each trial. For this response, children were required to respond until they answered correctly, with incorrect responses penalized by a buzzing noise and correct responses rewarded by a trumpet sound. The car then drove behind a single screen, making its pattern (but not wheels) occluded (Fig 1b). It jiggled behind the screen to indicate that a transformation was taking place. Children then responded with what they expected the car to look like when the screen lifted.

The experiment began with 6 explicit training trials in which children were provided with verbal feedback and instruction on the operation provided by each of the screens. Participants then entered an second training phase in which they received no verbal feedback from the experimenter, but were required to answer single-screen trials until correct, at which time the screen lifted to reveal the correct outcome (Fig 1c). Presentation of operations in this phase was in blocks of 6 single operations (red, blue, dots, stars, and two with no changes, in random order).

Children stayed in the training phase until they answered 5 out of 6 correct in a block. After meeting this criterion, the experiment progressed to a test phase. Here, children were shown blocks containing 2 single screen displays with feedback, and 4 two-screen displays *without* feedback (Fig 1d). In these critical trials, children observed the car pass behind one screen, then the next, without seeing it in between. As in training, they were required to answer correctly to the first question on the car’s initial pattern, but crucially received *no* feedback on their selected outcome after the car had passed behind both screens. The screens never lifted to reveal the car after selection and the experiment simply progressed to the next trial.

Screens were chosen to be iconic of the color and pattern transformations because we are primarily interested in how children combine these operations *when they know them*, not how hard it is for children to learn each individual function. However, sometimes the screen would not cause a change—for instance, if a car with blue dots drove behind a blue screen. These trials were included in order to ensure that children really paid attention to the pattern on the screens and did not just “flip” the relevant dimension. We refer to transformations in which a feature changed (e.g. red car behind blue screen) as *change* (CH) screens, and transformations which did not change a feature (e.g. blue car behind blue screen) as *identity* (ID) screens.

Children were run in the experiment for a maximum of approximately 30 minutes, and were shown a short “reward” animation every three responses in order to keep the experiment interesting. Stimuli were presented using Kelpy, the Kid Experimental Library in Python,

Fig 1. The four phases of the experiment. First, a truck is observed (a) and children are required to touch which pattern of the four possible patterns in the box matches the car. The response of the first pattern on the car stays on the screen (right box in (b)) and the car moves behind a screen with an iconic representation of its operation. Children are then asked to predict the outcome with a new set of four choices on the left (b). In training trials (b), children answer until they are correct and then see the screen lift to reveal the car with the new pattern (c). The critical response in test trials is shown in (d), with the car occluded after passing behind two screens. No feedback is provided on these trials.

doi:10.1371/journal.pone.0147734.g001

which is available under the GNU Public License from the first author [35]. This study was approved by the University of Rochester Research Subjects Review Board. Written informed consent was gathered from parents/guardians of the children involved.

2.1 Participants

Twenty-one children (10 females and 11 males) aged approximately 3.5–4.5 years (mean: 50.9 months; range: 42.9 to 53.9 months) were recruited to the Rochester Baby Lab. In order to assess overall levels of performance, all subjects were included in the following analysis, although two did not progress out of training.

3 Results

Fig 2 shows mean first response accuracies to single screen trials, broken down by each type of function: change (CH) color, change pattern, identity (ID) color, identity pattern. Means and standard errors were computed using a mixed effect logistic regression, taking into account

Fig 2. Mean response accuracies to single screens which performed change (CH) or identity (ID) operations. A chance rate of 25% is shown by the dotted red line.

doi:10.1371/journal.pone.0147734.g002

subject intercepts and slopes by function type [36]. This type of analysis is well-suited to unbalanced designs where children have been run for varying amounts of test and training items. The red line corresponds to a chance rate of 25%, for guessing at random from the four possible choices. This figure shows that in each condition children are substantially above chance on single screen displays. They thus learned the operations performed by each iconic screen. However, children are not very good at this task—their overall accuracy is near 50%. This likely results from the fact that the task is somewhat complex, requiring tracking of multiple values of multiple dimensions.

Critical two-screen test trials are shown in Fig 3. Again, error bars were computed via a mixed effect logistic regression with subject intercepts by function type (a model with random slopes did not converge). First, this figure shows that the overall mean response (teal bar) is substantially higher than the chance rate of 25%. The overall accuracy is also only slightly worse than the overall accuracy on single screen displays, meaning that knowledge of compositions is available to learners after training on single screens despite the fact that feedback was never provided on these two-screen trials.

Fig 3. Mean response accuracies to double screens, each of which performed change (CH) or identity (ID) operations. Error bars show confidence intervals. A chance rate of 25% is shown by the dotted red line. The blue dots correspond to the accuracy predicted under independent application of the single function accuracies in Fig 2.

doi:10.1371/journal.pone.0147734.g003

The other bars in this plot show performance broken down by the type of function performed by each of the two screens. Despite above chance performance overall, children are at chance in conditions involving feature changes within the same dimension (e.g. CH-Color/CH-Color). In such a condition, a red circle car would go behind a blue screen (changing its color) and then behind a red screen (changing its color again). This requires keeping track of several different values within the same dimension, and—because the features are binary—realizing that a feature changes and then reverts back to its original value. Note that children are not just having difficulty in this condition in tracking the order in which the functions applied: that would predict 50% performance in these conditions (since children would have the other dimension right). Instead, it appears that children's performance breaks down completely in this case, not unlike total breakdowns seen in object tracking [32–34].

These detailed patterns of responses also rule out several other hypotheses about children's performance. First, it is clear that children are not above chance by merely applying one of the two screens. Such a tactic would predict 0% accuracy on conditions with two changes (e.g. CH-Color-CH-Pattern and CH-Pattern-CH-Shape) since application of only one screen would *always* give the wrong answer. Instead, children performed particularly well in these two conditions, with a mean accuracy of nearly 50%, trending above even the overall experiment mean. Responses in these conditions are each individually significantly above the 25% chance level (CH-Color-CH-Pattern accuracy is 50%, $p = 0.003$ when compared to chance; CH-Pattern-

CH-Color is 55%, $p = 0.002$ when compared to chance). Children's success in these conditions provides strong evidence that they do not simply choose one box to apply.

Moreover, children's performance surpasses their expected performance based on their ability to apply single screens. The blue dots in Fig 3 show the performance that would be expected if children's accuracy was determined by independent application of each function, with individual function accuracies determined by the performance on single boxes. Thus, the blue dot for CH-Color/CH-Pattern corresponds to the probability of success on a CH-Color operation followed by the probability of success on a CH-Pattern operation, according to Fig 2. Because the accuracies in Fig 2 are near 50% for each function type, application of two of these correctly is near 25% ($= 0.5 \cdot 0.5$). Across nearly all function types other than changes within the same dimension, children are substantially above this performance level. This suggests that their failures are not due to independent failures on each screen. This pattern could occur if, for instance, children's low accuracy in Fig 2 was driven by not paying attention on all trials, rather than not knowing the right answer.

Responses were also analyzed using a mixed-effect logistic regression [36] with intercepts by subject (a model with random slopes did not converge). This analysis lets us simultaneously evaluate the influence of multiple factors of children's response accuracies, and determine their mean accuracy while controlling these factors. Fig 4 shows estimated coefficients and standard errors. The coefficients here correspond to whether the first function is an identity function (Identity F1), the second is an identity (Identity F2; this is computed as whether the second function is an identity function on the *output* of the first function), the first change is a color dimension (IsColor F1), the second is the same feature dimension as the first (SameDimension), children's age, whether the child is male, the child's training accuracy, the number of items it took the child to progress on to testing, and the number of test items seen so far. The binary predictors here (Identity F1, Identity F2, IsColor F1, SameDimension, IsMale) are all negative sum coded and the continuous predictors are all standardized, meaning that the intercept can be interpreted as the mean response accuracy across all predictors. In this figure, coefficients far from the zero line indicate significant influences on response accuracy. This shows that there are three significant predictors: children are substantially worse when both boxes operate on the same dimension (SameDimension) ($\beta = -0.50, z = -4.23, p < 0.001$). This means that when the first box changes a color, children are worse when the second box also changes a color, and analogously when both boxes change pattern. Children are worse when they take longer to reach testing ($\beta = -0.25, z = 2.14, p = 0.03$). Children with better training accuracy also perform better on testing ($\beta = 0.54, z = 4.40, p < 0.001$). Importantly, the intercept here represents the mean accuracy controlling for all other effects. This shows that the intercept ($\beta = -0.19$) is significantly higher than the chance rate $\text{logit}(0.25)$ shown by the red dot in the graph ($p < 0.001$), indicating that children are substantially above the chance guessing rate. They are not, however, significantly different from 50% accuracy ($p = 0.15$). These results revealed no significant effect of age, indicating that we have no evidence older children are better at this task. This might suggest that such compositional reasoning is attained earlier than about 3.5 years of age so that children in the range 3.5 ~ 4.5 are not differentially capable of dealing with composition.

We additionally analyzed the pattern of errors children made in their responses. To do this, we looked only at incorrect responses and tried to predict which components of an incorrect response children were likely to get correct. In a mixed effect regression, we found that these responses were substantially more likely to provide a response with the *second* screen correctly applied rather than the first ($\beta = 1.15, z = 4.946, p < 0.001$). This effect was independent of whether the first and second screens were colors or not ($|\beta| < 0.1, z < 0.5, p > 0.65$). This suggests that children's difficulty with composition may be in tracking or encoding the features of the car and the first screen while focusing on the second, temporally more recent, screen.

Fig 4. Coefficients in a mixed-effect logistic regression predicting accuracy on test (two-screen) trials from a number of predictors. There intercept here, representing the mean accuracy, should be compared to the chance rate of 1/4 (red dot, at $\text{logit}(0.25)$ on this scale), the overall chance guessing rate for 4 options. All other coefficients should be compared to the $x = 0$ line, showing whether they had a statistically significant influence on response accuracy. Effects significant at $p < 0.05$ are shown with gray bars.

doi:10.1371/journal.pone.0147734.g004

4 Discussion

These results provide evidence that preschoolers spontaneously infer the outcome of combinations of function applications after receiving training on only the individual pieces. Their performance on two screen displays—though not perfect—is comparable to their performance on single screen displays. These results suggest that knowledge of composition does not require training for preschoolers beyond learning the individual pieces: children at this age have already learned how to predictively combine operations. This does not mean that children have never required instruction (or data) for learning about composition, but it does mean that whatever they have learned before the experiment began is abstract enough to apply to novel functions.

It is worth noting that while our experiments were motivated by compositionality, there are likely *non*-compositional models that could capture these results. For instance, children may learn that each screen constrains the possible outcomes, and combine constraints in order to predict the outcome. The challenge for these theories would be to state combination in a way that cannot be viewed (or perhaps cannot be viewed naturally) as function composition. There

are also challenging data points for such theories even in this simple experiment: for instance, they would have a difficult time explaining why children are *below* 50% on within-dimension CH/CH trials, a fact that almost certainly depends on the details of how constraints may be combined. Stronger results could likely be provided by displays with more features and operations in order to fully probe the extent of children's ability and distinguish alternative models. We therefore take our results as a first step that provides suggestive—not definitive—evidence for compositionality.

Our analysis revealed some subtle facts about which compositions are easy and difficult for young learners. Changes within the same feature dimension are difficult, but two changes across dimensions are as easy as one. The limitation within dimensions presents a challenge to the purest form of compositional theories: what type of mechanism could correctly compose, but only when the composed functions operate on separate feature dimensions? One theory is that setting a feature value in memory is slow. Thus, when the value of two features change rapidly, children may get confused about what the resulting value is. On the other hand, if one feature changes and then a different feature changes, both “write” processes can occur in parallel and not interfere with each other. Difficulties dealing with rapidly sequential updates to a variable are common in parallel programming in computer science, (resulting in a so-called *race condition*). This view is consistent with the pattern of children's errors, where it appears that the transformation of the first screen is most likely to get overwritten or lost.

Interestingly, however, children's performance is also *below* 50%, indicating that their failure is perhaps more than a problem with remembering both screens. It is more likely to be a complete breakdown, because forgetting a single screen would still give 50% performance in the two-screen trials (since half the time, one will be an identity screen). This may indicate that memory mechanisms are fundamentally incapable of tracking multiple updates to the same feature dimension, perhaps because two rapid updates interfere in a particularly destructive way.

In the introduction, we discussed two forms of compositionality corresponding to whether learners explicitly represent a combination of functions $f_2 \circ f_1$ or whether they merely have the capacity to successively apply them to a representation $f_2(f_1(x))$. The present experiment does not strongly distinguish between these possibilities; it is possible that children only track the object moving behind the screens, successively updating its visual features (corresponding to $f_2(f_1(x))$). It is also possible that children could explicitly know that two screens together can be chunked into a single unit, the function $f_2 \circ f_1$. These possibilities might be disentangled by future work where children's ability to treat compositional functions as single units (e.g. in other compositions) could be evaluated.

In addition to demonstrating compositional ability, these results also suggest that children aged 3.5 ~ 4.5 are able to represent *functions*, not itself a trivial capacity. In particular, to perform above chance children must be able to represent the fact that each of four screens performs a particular change to an object's features, and that that change occurs even if the outcome is not directly observed. Such an ability to represent functions themselves might be viewed as an even more basic ability than compositionality. However, fluency with functions can only yield more complex computations if those functions can be combined to form novel combinations—which our results indicate preschoolers are likely able to do.

We consider this capacity for representing functions themselves as potentially a basic fact about the organization of cognitive computation. In modern computer systems, encapsulated functions play a critical role by allowing complex computations to efficiently and easily be built from simpler components. For instance, programs are constructed only out of simpler elementary operations (e.g. +, −, ·, /, *for*, *if*, etc.) which are eventually directly interpretable by the computer's hardware. Using only these kinds of primitive functions, one can create a structure

implementing a more complex computation, like $f(x) = x \cdot x \cdot x + x - (x+1)/x$. The capacity to combine such elementary operations in arbitrary ways is extremely powerful, and an ability to manipulate functions themselves potentially allows a huge range of computations to be executed using very little “built in” knowledge (ie. few primitives). Indeed, the simplest computational systems like lambda calculus [23, 25] and combinatory logic [24, 37] “build in” almost nothing, essentially *only* the rules of function composition. A striking result in formal logic holds that arbitrary (ie. Turing-complete) computational processes can then be built out of nothing more than this capacity for composition [23]. This means that the ability of children to represent, manipulate, and compose functions may point to how they are able to acquire operations of substantial computational complexity, while requiring only very simple cognitive machinery. In this sense, compositionality may be *the* key component of LOT theories that allows arbitrarily complex computations to be represented by learners.

Our study was in large part motivated by LOT learning models, but it is also interesting to consider how to interpret these results in the context of larger theoretical divides in cognitive science. It may be productive, for example, to determine how such learning might be captured in connectionist or parallel distribution processing [38] approaches: how might one capture learning of separate functions which can be composed, without requiring any additional training on composition? Is it possible to set up such a system in a way that cannot be interpreted as presupposing compositionality itself? Or, it may be that compositionality is basic enough that, as in LOT theories, it should be assumed as a core computational primitive.

5 Conclusion

These results are an early step in linking contemporary structured learning models with infant and early childhood experimental studies. An ability to combine novel functions appears to be robustly present by three or four years of age, with children requiring no explicit training on combination once they have learned individual functions. This suggests that learning theories based around assuming compositionality are behaviorally plausible and that a capacity for combining mental operations may be one of the mechanisms that supports children’s creation of rich conceptual systems.

Acknowledgments

This work benefited greatly from discussions with Roman Feiman, Celeste Kidd, Noah Goodman, and members of AKlab. Children were run with the help of Holly Palmeri and the Rochester Babylab RAs. Research reported in this publication was supported by the Eunice Kennedy Shriver National Institute Of Child Health & Human Development of the National Institutes of Health under Award Number F32HD070544 and an NIH research grant to R. Aslin and E. Newport (HD-037082). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Author Contributions

Conceived and designed the experiments: SP RA.

Performed the experiments: SP.

Analyzed the data: SP.

Wrote the paper: SP RA.

References

1. Rosch E, Mervis CB. Family resemblances: Studies in the internal structure of categories. *Cognitive psychology*. 1975; 7(4):573–605. doi: [10.1016/0010-0285\(75\)90024-9](https://doi.org/10.1016/0010-0285(75)90024-9)
2. Medin DL, Schaffer MM. Context theory of classification learning. *Psychological review*. 1978; 85(3):207–238. doi: [10.1037/0033-295X.85.3.207](https://doi.org/10.1037/0033-295X.85.3.207)
3. Smith EE, Medin DL. The exemplar view. *Foundations of cognitive psychology: Core readings*. 2002; p. 277–292.
4. Fodor JA. *The language of thought*. Cambridge, MA: Harvard University Press; 1975.
5. Margolis E., Laurence S. *Concepts: Core Readings*. Cambridge, MA: The MIT Press; 1999.
6. Osherson DN, Smith EE. On the adequacy of prototype theory as a theory of concepts. *Cognition*. 1981; 9(1):35–58. doi: [10.1016/0010-0277\(81\)90013-5](https://doi.org/10.1016/0010-0277(81)90013-5) PMID: [7196818](https://pubmed.ncbi.nlm.nih.gov/7196818/)
7. Smith EE, Osherson DN. Conceptual combination with prototype concepts. *Cognitive Science*. 1984; 8(4):337–361. doi: [10.1207/s15516709cog0804_2](https://doi.org/10.1207/s15516709cog0804_2)
8. Kamp H, Partee B. Prototype theory and compositionality. *Cognition*. 1995; 57(2):129–191. doi: [10.1016/0010-0277\(94\)00659-9](https://doi.org/10.1016/0010-0277(94)00659-9) PMID: [8556840](https://pubmed.ncbi.nlm.nih.gov/8556840/)
9. Fodor J, Lepore E. The red herring and the pet fish: Why concepts still can't be prototypes. *Cognition*. 1996; 58(2):253–270. doi: [10.1016/0010-0277\(95\)00694-X](https://doi.org/10.1016/0010-0277(95)00694-X) PMID: [8820389](https://pubmed.ncbi.nlm.nih.gov/8820389/)
10. Smith EE, Osherson DN, Rips L, Keane M. Combining prototypes: A selective modification model. *Cognitive Science*. 1988; 12(4):485–527. doi: [10.1207/s15516709cog1204_1](https://doi.org/10.1207/s15516709cog1204_1)
11. Fodor J, Pylyshyn ZW. Connectionism and cognitive architecture: a critical analysis, *Connections and symbols*. A Cognition Special Issue, Pinker S and Mehler J (eds). 1988; p. 3–71.
12. Boole G. *An investigation of the laws of thought: on which are founded the mathematical theories of logic and probabilities*. London, UK: Walton and Maberly; 1854.
13. Frege G. Über sinn und bedeutung. *Wittgenstein Studien*. 1892; 1(1).
14. Siskind JM. A Computational Study of Cross-Situational Techniques for Learning Word-to-Meaning Mappings. *Cognition*. 1996; 61:31–91. doi: [10.1016/S0010-0277\(96\)00728-7](https://doi.org/10.1016/S0010-0277(96)00728-7)
15. Katz Y, Goodman ND, Kersting K, Kemp C, Tenenbaum JB. Modeling semantic cognition as logical dimensionality reduction. In: *Proceedings of Thirtieth Annual Meeting of the Cognitive Science Society*; 2008.
16. Goodman ND, Tenenbaum JB, Feldman J, Griffiths TL. A Rational Analysis of Rule-Based Concept Learning. *Cognitive Science*. 2008; 32(1):108–154. doi: [10.1080/03640210701802071](https://doi.org/10.1080/03640210701802071) PMID: [21635333](https://pubmed.ncbi.nlm.nih.gov/21635333/)
17. Kemp C, Goodman ND, Tenenbaum JB. Learning and using relational theories. *Advances in neural information processing systems*. 2008; 20:753–760.
18. Goodman ND, Ullman TD, Tenenbaum JB. Learning a theory of causality. In: *Proceedings of the 31st annual conference of the cognitive science society*; 2009. p. 2188–2193.
19. Ullman TD, Goodman ND, Tenenbaum JB. Theory Acquisition as Stochastic Search. In: *Proceedings of thirty second annual meeting of the cognitive science society*; 2010.
20. Piantadosi ST. *Learning and the language of thought*. MIT; 2011. Available from: http://colala.bcs.rochester.edu/papers/piantadosi_thesis.pdf.
21. Piantadosi ST, Tenenbaum JB, Goodman ND. Bootstrapping in a language of thought: a formal model of numerical concept learning. *Cognition*. 2012; 123:199–217. doi: [10.1016/j.cognition.2011.11.005](https://doi.org/10.1016/j.cognition.2011.11.005) PMID: [22284806](https://pubmed.ncbi.nlm.nih.gov/22284806/)
22. Feldman J. Minimization of Boolean complexity in human concept learning. *Nature*. 2000; 407(6804):630–633. doi: [10.1038/35036586](https://doi.org/10.1038/35036586) PMID: [11034211](https://pubmed.ncbi.nlm.nih.gov/11034211/)
23. Church A. An unsolvable problem of elementary number theory. *American journal of mathematics*. 1936; 58(2):345–363. doi: [10.2307/2371045](https://doi.org/10.2307/2371045)
24. Schönfinkel M. On the building blocks of mathematical logic. From Frege to Gödel. 1967; p. 355–366.
25. Hindley JR, Seldin JP. *Introduction to Combinators and λ -calculus*. Cambridge, UK: Press Syndicate of the University of Cambridge; 1986.
26. Matthei EH. The acquisition of prenominal modifier sequences. *Cognition*. 1982; 11(3):301–332. doi: [10.1016/0010-0277\(82\)90018-X](https://doi.org/10.1016/0010-0277(82)90018-X) PMID: [7199414](https://pubmed.ncbi.nlm.nih.gov/7199414/)
27. Barner D, Snedeker J. Compositionality and Statistics in Adjective Acquisition: 4-Year-Olds Interpret Tall and Short Based on the Size Distributions of Novel Noun Referents. *Child development*. 2008; 79(3):594–608. doi: [10.1111/j.1467-8624.2008.01145.x](https://doi.org/10.1111/j.1467-8624.2008.01145.x) PMID: [18489415](https://pubmed.ncbi.nlm.nih.gov/18489415/)

28. Fernald A, Thorpe K, Marchman VA. Blue car, red car: Developing efficiency in online interpretation of adjective–noun phrases. *Cognitive psychology*. 2010; 60(3):190–217. doi: [10.1016/j.cogpsych.2009.12.002](https://doi.org/10.1016/j.cogpsych.2009.12.002) PMID: [20189552](https://pubmed.ncbi.nlm.nih.gov/20189552/)
29. Syrett K, Lidz J. 30-month-olds use the distribution and meaning of adverbs to interpret novel adjectives. *Language Learning and Development*. 2010; 6(4):258–282. doi: [10.1080/15475440903507905](https://doi.org/10.1080/15475440903507905)
30. Wimmer H, Perner J. Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*. 1983; 13(1):103–128. doi: [10.1016/0010-0277\(83\)90004-5](https://doi.org/10.1016/0010-0277(83)90004-5) PMID: [6681741](https://pubmed.ncbi.nlm.nih.gov/6681741/)
31. Wynn K. Addition and subtraction by human infants. *Nature*. 1992; 358(6389):749–750. doi: [10.1038/358749a0](https://doi.org/10.1038/358749a0) PMID: [1508269](https://pubmed.ncbi.nlm.nih.gov/1508269/)
32. Baillargeon R, Miller K, Constantino J. Ten-month-old infants' intuitions about addition. Unpublished manuscript, University of Illinois at Urbana, Champaign, IL. 1994;.
33. Feigenson L, Carey S. On the limits of infants' quantification of small object arrays. *Cognition*. 2005; 97(3):295–313. doi: [10.1016/j.cognition.2004.09.010](https://doi.org/10.1016/j.cognition.2004.09.010) PMID: [16260263](https://pubmed.ncbi.nlm.nih.gov/16260263/)
34. Moher M, Tuerk AS, Feigenson L. Seven-month old infants chunk items in working memory. *Journal of Experimental Child Psychology*. 2012; 112:361–377. doi: [10.1016/j.jecp.2012.03.007](https://doi.org/10.1016/j.jecp.2012.03.007) PMID: [22575845](https://pubmed.ncbi.nlm.nih.gov/22575845/)
35. Piantadosi ST. Kelpy: a free library for child experimentation in python; 2012. available from <https://github.com/piantado/kelpy/>.
36. Gelman A, Hill J. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge, UK: Cambridge University Press; 2007.
37. Smullyan RM. *To Mock a Mockingbird: and other logic puzzles including an amazing adventure in combinatory logic*. Oxford University Press; 1985.
38. Rumelhart DE, McClelland JL. *Parallel distributed processing*. Cambridge, MA: MIT Press; 1986.