Association for Psychological Science

Current Directions in Psychological Science
2016, Vol. 25(1) 54–59

DOI: 10.1177/0963721415609581
cdps.sagepub.com
SAGE

Four Problems Solved by the Probabilistic Language of Thought

Steven T. Piantadosi and Robert A. Jacobs
Department of Brain and Cognitive Sciences, University of Rochester

Abstract
We argue for the advantages of the probabilistic language of thought (pLOT), a recently emerging approach to modeling human cognition. Work using this framework demonstrates how the pLOT (a) refines the debate between symbols and statistics in cognitive modeling, (b) permits theories that draw on insights from both nativist and empiricist approaches, (c) explains the origins of novel and complex computational concepts, and (d) provides a framework for abstraction that can link sensation and conception. In each of these areas, the pLOT provides a productive middle ground between historical divides in cognitive psychology, pointing to a promising way forward for the field.

Keywords
language of thought, modeling

The use of structured symbolic representations in cognitive theories has a long history in cognitive psychology, dating back to Boole's (1854) development of symbolic logic to characterize the "laws of thought." The approach reached prominence through the work of Fodor, who argues for the existence of a language-like system of mental representations, a *language of thought* (LOT; Fodor, 1975, 2008). In this formalism, concepts are formed through structured compositions of symbols. For instance, a mental representation of *uncle* (as a relation between two individuals) might be realized as a logical structure,

$$UNCLE(x, y) := \exists z.\, SIBLING(x, z) \wedge PARENT(z, y). \quad (1)$$

Here, *uncle* is defined in terms of other—likely simpler—operations, including *sibling, parent,* existential quantification (∃), and logical conjunction (∧), with free variables (x, y, and z) representing the individuals involved. These operations may in turn be composed of even simpler concepts, but eventually all operations reduce to primitives that are assumed to be innate. In this way, the LOT shares much of its spirit with programming languages, whose power comes from an ability to combine a small number of built-in operations to express unboundedly complex ideas. The LOT naturally handles problems like the systematicity, productivity, and compositionality of thinking (Fodor & Pylyshyn, 1988), and it has the advantage of explaining cognitive processes via symbols that are themselves interpretable and testable from the outside.

In its basic form, the LOT represents a claim about representation, not learning or inference. However, a number of learning theories have drawn on LOT formalisms and attempted to explain how learners might decide which compositions of primitives they should construct in the face of data. For instance, the classic studies of Bruner, Goodnow, and Austin (1956) examined the discovery of LOT-like Boolean rules in simple domains (see also Hunt, Marin, & Stone, 1966). A concept might be

$$red(x) \wedge \left(square(x) \vee circle(x)\right), \quad (2)$$

where subjects would look at data (e.g., positive and negative instances of objects with shape and color features) and try to deduce the correct particular combination of features (e.g., *red, square, circle*) and logical connectives (≠, ∧, and ∨). Bruner et al. provided heuristic or deductive theories of human performance based on falsification of rules by examples. Siskind (1996) similarly studied the

Corresponding Author:
Steven Piantadosi, Department of Brain and Cognitive Sciences, University of Rochester, 358 Meliora Hall, Box 270268, Rochester, NY 14627
E-mail: spiantado@gmail.com

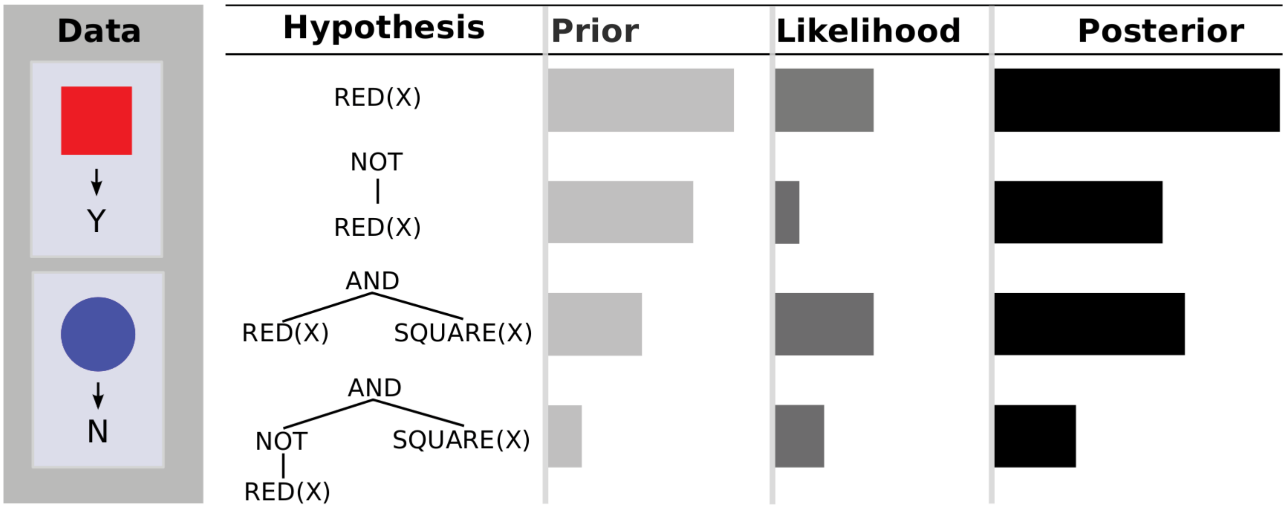

Fig. 1. In the probabilistic language of thought (pLOT), hypotheses are constructed by composing a small set of primitive operations according to a grammar. Simple compositions are assigned higher priors, which are then combined with a statistical likelihood measuring how well each hypothesis explains the data—here, mappings from shapes to Boolean values (Y = yes, N = no) that are assumed to be noisy. Hypotheses that achieve a trade-off between simplicity (prior) and fit to data (likelihood) are found to have the highest posterior probability, reflecting rational learners' theorized beliefs.

acquisition of word meanings in a compositional representation language, providing a heuristic algorithm for building compositional word meanings.

More recently, the LOT has been combined with inferentially rigorous systems that model probabilistic beliefs, drawing motivation from both ideal-observer analyses and Bayesian cognitive theories more generally (e.g., Geisler, 2003; Tenenbaum, 1999). The first such example was Goodman, Tenenbaum, Feldman, and Griffiths (2008), who showed how probabilistic inference could be used to model people's acquisition of Boolean concepts, such as that shown in (2). Building on Feldman's (2000) findings that logical simplicity influences learning, Goodman et al.'s model combines a prior favoring rule simplicity with a noisy likelihood function via Bayesian inference.[1] This work showed that phenomena in Boolean concept learning, like selective-attention effects and simplicity preferences, could be understood as results from rationally determining how to combine logical primitives into concepts.

This model can be viewed as the first formulation of the *probabilistic* language of thought (pLOT), combining structured, rule-like representations with Bayesian probabilistic inference to solve problems of learning and inference. The logic of the pLOT is illustrated in Figure 1, in which Bayesian inference is used to decide how likely any particular composition of functions (rows) is, given some observed data. The inductive use of a LOT builds on other work in artificial intelligence (e.g., De Raedt, 2008; Richardson & Domingos, 2006) and psychology (e.g., Anderson, 1996) that has married rules with probability and inference. Goodman, Tenenbaum, and Gerstenberg (2015) described a version of the pLOT in which the evaluation of primitives themselves is stochastic, meaning that a single expression yields a distribution of values. By conditioning on one of these values, their inference engines are able to discover what random choices were made in the evaluation of a (stochastic) pLOT expression, a framework implemented in the programming language Church (Goodman, Mansinghka, Roy, Bonawitz, & Tenenbaum, 2008). They argue that this marriage of rules, probability, and Bayesian inference provides a means for unifying rule-like, fuzzy, and compositional aspects of human concepts. Recently, pLOT models have been proposed in a variety of domains, bringing into focus a new unifying perspective on a number of central debates in cognitive psychology.

Symbols and Statistics

Historically, there have been two primary frameworks for cognitive modeling. One school of thought favors symbolic approaches, typically based on grammars, production rules, or logic. These approaches focus on the processes for combining and manipulating symbols to form new symbols (often of a more abstract nature) and complex multisymbol expressions and structures. This permits rich expressiveness via compositionality but is often criticized for being "brittle," failing to adequately characterize domains with inherently noisy or ambiguous input. In contrast, statistical approaches (e.g., Bayesian networks) are flexible and robust, able to handle ambiguity and uncertainty in a principled way. Of course, their flexibility and adaptability come at a price. These approaches often require highly structured ways of characterizing prior beliefs (e.g., in the form of prior distributions) and new experiences (in the form of likelihood functions).

Since the 1950s, the dominance of either symbolic or statistical schools of thought have waxed and waned. At several moments, advocates of these different frameworks have engaged in heated debates engaging the question of rules versus statistics as well as the appropriate cognitive architecture (McClelland & Patterson, 2002a, 2002b; Pinker & Ullman, 2002a, 2002b). Prior debates have not led to a consensus as to which approach is best—and probably for good reason. The cognitive system is probably both symbolic and statistical. Indeed, Feldman (2012) developed a unifying view of symbols as arising from effective representations of environments with sparse components. In his formalism, symbols are justified only under some assumptions, potentially providing a way to understand why humans have both symbolic and nonsymbolic representations. The pLOT takes the advantages of both symbols and statistics in order to formalize a hybrid middle ground. Like symbolic approaches, pLOT systems use symbols and processes for combining and manipulating symbols. Consequently, these systems are capable of generating rich representations. However, the application of processes to symbols is probabilistic, not deterministic. The pLOT uses probabilistic inference to learn from data which processes should be applied to which symbols. Because this learning is statistical, it can be robust to noise and ambiguity.

A prototypical domain for this problem is word learning, in which children must take observations of words across contexts and infer their underlying meaning. Typically, a single context will not fully disambiguate word meaning; at the same time, many word meanings are abstract and logical. Function words (e.g., *many, the, of, and*, etc.), for instance, do not refer to observable properties in the world but rather determine how other words in a sentence combine. Building on Siskind (1996), recent pLOT work in the case of quantifiers (Piantadosi, Goodman, & Tenenbaum, 2015) and number words (Piantadosi, Tenenbaum, & Goodman, 2012) showed how the pLOT could learn these meanings in a way that adequately handled noise, ambiguous evidence, and abstract logical semantics.

Nativism and Empiricism

Though it is rarely described as such, the historical debate between nativism and empiricism in cognitive psychology has been extremely fruitful. In the prototypical domain of language acquisition, both sides have toed the boundaries of plausible theories, allowing the field to understand the positives and negatives of a wide range of assumptions—from the theory that learners "build in" only architectural constraints to the theory that they essentially come with a full-fledged grammar. The main lesson from this debate is that both sides are deeply unsatisfying. The strongest empiricist versions of language learning fail to provide workable theories of learning that are connected to children's incremental performance and adults' remarkable ability. Strongly nativist theories have neglected the power of modern machine learning, instead falling back on what has been called an "argument from lack of imagination" about how learning could possibly succeed.[2]

Children clearly have something special—something lacking in other animals—that permits language acquisition by humans alone. At the same time, children do *learn* something substantial about how language works, given the remarkable diversity of languages children acquire (Evans & Levinson, 2009). These facts leave us in need of a theoretical approach that can integrate the built-in capacities future work will discover with whatever genuine learning and inference we observe.

The pLOT provides an appealing approach to this end because it necessarily has both innate and learned components, and these are made explicit in any model. The primitives are assumed to be either innate or learned at an earlier time during development, and the set of possible compositions of primitives form learners' innate space of hypotheses in any particular domain. For instance Piantadosi et al. (2012) built in primitives corresponding to a hypothesis about children's "core knowledge" (Carey, 2009; Spelke, 2003; Spelke & Kinzler, 2007) in the domain of number, including an ability to represent sets and perform basic set-theoretic operations. The choice of innate primitives can be viewed as a strictly empirical question that should be determined through independent experiments. At the same time, the model also captured genuine learning, as these abilities on their own were not the same as knowledge of number and counting. This provided a concrete (implemented) demonstration of the sense in which number might be constructed (Xu & Kushnir, 2013) out of children's innate repertoire. The explicit differentiation between what is built in and what is learned permits the pLOT to tread the middle grounds between nativist and empiricist extremes.[3]

Novelty in Learning

One outstanding mystery of cognition is where novel complex concepts may come from. Adults clearly possess a large number of complex concepts that children lack, yet there is no theory of how such rich knowledge and computation might arise. This is most striking when we consider the range of things ordinary adults do—things like reasoning through what likely caused car troubles, determining which ingredients in a recipe can be altered, or guessing which present a relative will like most. Each of these is an extremely complex computation, drawing on rich types of world knowledge, deduction, induction, and knowledge of intricate causal mechanisms. How can we start with the cognitive mechanisms infants possess and arrive at such rich cognitive processes?

Fodor (1975) has developed the case for "extreme nativism," in which even our complex concepts—most famously the concept of a carburetor—are innate primitives. Fodor's reasoning is that most concepts are not compositional; however, in LOT learning, all that *can* be learned is compositions of primitive operations. If most concepts are not compositional, but only compositions can be learned, then the only way these complex concepts (e.g., *carburetor*) can get into cognitive systems is as innate primitives.

Fodor's argument is a curious one to make for someone who is otherwise such a proponent of the computational theory of mind. The reason is that everything that can be computed can be characterized compositionally, a deep finding in mathematical logic. Computability is most famously formalized via Turing machines. However, contemporaries of Turing such as Church (1936) alternatively developed *lambda calculus*, a logical system for computation that is based entirely on function composition. A composition of functions evaluates ("runs") through function application to yield a given value (or does not halt evaluation). This approach to computation survives today in functional programming languages such as Scheme and Haskell and forms the foundation for the programming language Church (Goodman, Mansinghka, et al., 2008). Computability for lambda calculus is equivalent to Turing computability (Turing, 1937), meaning that any computation in one can be translated to the other. The existence of a formalism that is (a) Turing-complete and (b) based entirely on compositions nullifies Fodor's argument. If one believes a computational theory of mind, then everything the mind does—including compute the meaning of *carburetor*—can be expressed as compositions of logical operations. Fodor's argument for radical concept nativism fails because computability implies compositionality.

The view from pLOT is again a middle ground: Learners come with some special innate primitives—namely, a set of Turing-complete operators—and use these to build new representations. The formal requirements are extraordinarily minimal: Lambda calculus expresses the laws of function combination, and combinatory logic (see Hindley & Seldin, 1986) expresses all computations as compositions of a few higher-order functions. In theory, one can build in just these Turing-complete minimal systems, allowing learners to acquire new representations as compositions. When this representational capacity is combined with the pLOT's inductive machinery, learners could in principle acquire representations of unbounded computational complexity. This theoretical perspective points to what we believe to be the only formal system currently capable—even in principle—of acquiring the arbitrarily complex computational representations adults possess.

From Sensation to Conception

In the domain of perception, it seems clear that people have multiple representations of objects and events. Some representations are more sensory in nature, whereas others are more conceptual. This is illustrated by crossmodal transfer of knowledge: If a person is trained to visually categorize a set of objects, he or she will often be able to categorize novel objects from the same categories when those objects are grasped but not seen (Wallraven, Bülthoff, Waterkamp, van Dam, & Gaißert, 2014; Yildirim & Jacobs, 2013). This suggests that people possess conceptual representations that can characterize objects and events (e.g., object shape) in a modality-independent manner. How might modality-independent conceptual representations arise from modality-specific sensory representations?

The pLOT combines the necessary structure and robustness to address this question by simultaneosly describing domains at multiple levels of abstraction. The grammar behind a pLOT theory can describe how representations are built on multiple scales—for instance, in the domain of vision, it can specify how local patches combine to form surfaces, how surfaces combine to form parts, how parts combine to form objects, and how objects combine to form scenes. This is the main idea underlying computer graphics systems. In these systems, an object's shape, for instance, is described using a mesh of hundreds or thousands of small polygons. Because each polygon is small and simple, it is easy to map it to its visual features; because it is part of a larger schematic, it is easy to integrate it with high-level structures that determine, for instance, its position relative to other patches.

Yildirim and Jacobs (2015) showed concretely how pLOT models can bridge sensation to conception in an environment with both visual and auditory sequences of spatial locations. Their system included three main components: (a) a probabilistic grammar that characterized spatial sequences in a modality-independent manner (e.g., "move one unit clockwise," "move two units counterclockwise," etc.); (b) sensory-specific forward models that mapped modality-independent sequence representations to corresponding visual (perhaps a form of visual imagery) or auditory (perhaps a form of auditory imagery) features; and (c) a Bayesian-inference algorithm that inferred modality-independent sequence representations from either visual or auditory features despite significant sensory noise in these features. Because this system contained both sensory and conceptual representations of sequences, it showed successful crossmodal transfer of knowledge—after training to categorize sequences based on visual information, it was able to categorize these same sequences when they were heard but not seen, and vice versa. Yildirim and Jacobs (2013) and Erdogan, Yildirim, and Jacobs (2015) developed related systems containing both sensory and conceptual representations of objects when objects were viewed, grasped, or both. These systems showed crossmodal transfer of object-shape knowledge across visual and haptic modalities. The pLOT is unique in that it provides a single framework supporting representation and learning that can process everything from low-level feature-based primitives to abstract compositional functions.

Conclusion

The pLOT is not a revolutionary new theory that promises to overthrow existing paradigms; it is a resurgent old theory that promises to integrate many approaches into a unitary framework. However, the pLOT research program reviewed here will not, by itself, make the link to the biological systems supporting cognition. To do so, research on the pLOT must eventually connect to neural and cognitive theories of how symbolic LOTs can arise out of subsymbolic systems. Until then, we argue that it provides one of the most promising frameworks for cognition, combining the compositionality of symbolic approaches with the robustness of probabilistic approaches, thereby permitting researchers to formulate and test theories that do not acquiesce to the poles of major debates.

Recommended Reading

Erdogan, G., Yildirim, I., & Jacobs, R. A. (2015). (See References). Shows how a probabilistic LOT can explain the learning of representations that span individual modalities and sensory systems.

Feldman, J. (2000). (See References). Shows human sensitivity to logical complexity in learning tasks, thus providing evidence for the psychological reality of logical representations like those posited by the LOT.

Fodor, J. (1975). (See References). A philosophical overview of the arguments supporting a LOT.

Goodman, N. D., & Lassiter, D. (2015). Probabilistic semantics and pragmatics: Uncertainty in language and thought. In S. Lappin & C. Fox (Eds.), *The Handbook of Contemporary Semantic Theory* (2nd ed., pp. 655-686). Hoboken, NJ: Wiley-Blackwell. Applies a probabilistic language to understanding semantic and pragmatic theories and phenomena in linguistics.

Goodman, N., Tenenbaum, J., Feldman, J., & Griffiths, T. (2008). (See References). The first work to combine logical LOT representations with a Bayesian inferential model, which develops methods and techniques commonly used by later LOT models.

Piantadosi, S., Tenenbaum, J., & Goodman, N. (2012). (See References). Shows how children's patterns in learning number words can be explained by a probabilistic LOT, helping to direct the philosophical debate about whether concepts such as number are necessarily innate.

Acknowledgments

We are extremely grateful to Jay McClelland and Jacob Feldman for providing detailed and insightful comments on this article.

Declaration of Conflicting Interests

The authors declared that they had no conflicts of interest with respect to their authorship or the publication of this article.

Notes

1. This work also developed one of the first Monte Carlo inference algorithms capable of working with a LOT—a generalization of the Metropolis–Hastings algorithm that provided a sampler for LOT expressions given data.
2. Even the formal results, such as those of Gold (Gold, 1967), rely on the assumption of parents who are infinitely antagonistic to their children's success (see Johnson, 2004).
3. It also makes the pLOT itself testable: The pLOT would fail as a framework if one could provide strong evidence for a set of primitives (operations) that children can use but strong evidence against the developmental trajectory those primitives would predict under pLOT learning.

References

Anderson, J. (1996). ACT: A simple theory of complex cognition. *American Psychologist*, *51*, 355–365.

Boole, G. (1854). *An investigation of the laws of thought: On which are founded the mathematical theories of logic and probabilities*. London, England: Walton & Maberly.

Bruner, J., Goodnow, J., & Austin, G. (1956). *A study of thinking*. New Brunswick, NJ: Transaction Publishers.

Carey, S. (2009). *The origin of concepts*. Oxford, England: Oxford University Press.

Church, A. (1936). An unsolvable problem of elementary number theory. *American Journal of Mathematics*, *58*, 345–363.

De Raedt, L. (2008). *Logical and relational learning*. New York, NY: Springer-Verlag.

Erdogan, G., Yildirim, I., & Jacobs, R. A. (2015). From sensory signals to modality-independent conceptual representations: A probabilistic language of thought approach. *PLoS Computational Biology*, *11*(11), e1004610. doi:10.1371/journal.pcbi.1004610

Evans, N., & Levinson, S. (2009). The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral & Brain Sciences*, *32*, 429–448.

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

Feldman, J. (2012). Symbolic representation of probabilistic worlds. *Cognition*, *123*, 61–83.

Fodor, J. (1975). *The language of thought*. Cambridge, MA: Harvard University Press.

Fodor, J. (2008). *LOT 2: The language of thought revisited*. Oxford, England: Oxford University Press.

Fodor, J., & Pylyshyn, Z. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition*, *28*(1–2), 3–71.

Geisler, W. (2003). Ideal observer analysis. In L. M. Chalupa & J. S. Werner (Eds.) *The visual neurosciences* (pp. 825–837). Cambridge, MA: Bradford Books.

Gold, E. (1967). Language identification in the limit. *Information and Control*, *10*, 447–474.

Goodman, N., Mansinghka, V., Roy, D., Bonawitz, K., & Tenenbaum, J. (2008). Church: A language for generative models. In D. McAllester & P. Myllymaki (Eds.), *Proceedings of the 24th conference on uncertainty in artificial intelligence, UAI 2008* (pp. 220–229). Corvallis, OR: AUAI Press.

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

Goodman, N., Tenenbaum, J., & Gerstenberg, T. (2015). Concepts in a probabilistic language of thought. In Margolis & Laurence (Eds.), *The conceptual mind: New directions in the study of concepts* (pp. 623–654). Cambridge, MA: MIT Press.

Hindley, J., & Seldin, J. (1986). *Introduction to combinators and λ-calculus*. Cambridge, England: Press Syndicate of the University of Cambridge.

Hunt, E., Marin, J., & Stone, P. (1966). *Experiments in induction*. Oxford, England: Academic Press.

Johnson, K. (2004). Gold's theorem and cognitive science. *Philosophy of Science*, *71*, 571–592.

McClelland, J. L., & Patterson, K. (2002a). Rules or connections in past-tense inflections: What does the evidence rule out? *Trends in Cognitive Sciences*, *6*, 465–472.

McClelland, J. L., & Patterson, K. (2002b). 'Words or rules' cannot exploit the regularity in exceptions. *Trends in Cognitive Sciences*, *6*, 464–465.

Piantadosi, S., Goodman, N., & Tenenbaum, J. (2015). *Modeling the acquisition of quantifier semantics: A case study in function word learnability*. Manuscript submitted for publication.

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

Pinker, S., & Ullman, M. T. (2002a). Combination and structure, not gradedness, is the issue. *Trends in Cognitive Sciences*, *6*, 472–474.

Pinker, S., & Ullman, M. T. (2002b). The past and future of the past tense. *Trends in Cognitive Sciences*, *6*, 456–463.

Richardson, M., & Domingos, P. (2006). Markov logic networks. *Machine Learning*, *62*, 107–136.

Siskind, J. (1996). A computational study of cross-situational techniques for learning word-to-meaning mappings. *Cognition*, *61*, 31–91.

Spelke, E. (2003). What makes us smart? Core knowledge and natural language. In D. Gentner & S. Goldin-Meadow (Eds.), *Language in mind* (pp. 277–312). Cambridge, MA: MIT Press.

Spelke, E., & Kinzler, K. (2007). Core knowledge. *Developmental Science*, *10*, 89–96.

Tenenbaum, J. (1999). *A Bayesian framework for concept learning* (PhD thesis). Cambridge, MA: Massachusetts Institute of Technology.

Turing, A. M. (1937). Computability and λ-definability. *The Journal of Symbolic Logic*, *2*, 153–163.

Wallraven, C., Bülthoff, H. H., Waterkamp, S., van Dam, L., & Gaißert, N. (2014). The eyes grasp, the hands see: Metric category knowledge transfers between vision and touch. *Psychonomic Bulletin & Review*, *21*, 976–985.

Xu, F., & Kushnir, T. (2013). Infants are rational constructivist learners. *Current Directions in Psychological Science*, *22*, 28–32.

Yildirim, I., & Jacobs, R. A. (2013). Transfer of object category knowledge across visual and haptic modalities: Experimental and computational studies. *Cognition*, *126*, 135–148.

Yildirim, I., & Jacobs, R. A. (2015). Learning multisensory representations for auditory-visual transfer of sequence category knowledge: A probabilistic language of thought approach. *Psychonomic Bulletin & Review*, *22*, 673–686.