Modeling infant object perception as program induction

J.-Philipp Fränken[1*], **Christopher G. Lucas**[2], **Neil R. Bramley**[2], **Steven T. Piantadosi**[3]
[1]Stanford University, [2]The University of Edinburgh, [3]University of California, Berkeley, *jphilipp@stanford.edu

Abstract

Infants expect physical objects to be rigid and persist through space and time and in spite of occlusion. Developmentists frequently attribute these expectations to a "core system" for object recognition. However, it is unclear if this move is necessary. If object representations emerge reliably from general inductive learning mechanisms exposed to small amounts of environment data, it could be that infants simply induce these assumptions very early. Here, we demonstrate that a domain general learning system, previously used to model concept learning and language learning, can also induce models of these distinctive "core" properties of objects after exposure to a small number of examples. Across eight micro-worlds inspired by experiments from the developmental literature, our model generates concepts that capture core object properties, including rigidity and object persistence. Our findings suggest infant object perception may rely on a general cognitive process that creates models to maximize the likelihood of observations.[1]

Keywords: core knowledge; perception; vector quantization; program induction; Bayes

Introduction

Object representations serve as compositional building blocks for higher level cognition in both humans and machines (Xu & Carey, 1996; Schölkopf et al., 2021; Chen et al., 2022). Developmental accounts suggest that infants rely on a "core system" for object representations to perceive the boundaries of objects, accurately represent their shapes even when they are partially or fully occluded, and make predictions about object movements and their final positions (Spelke & Kinzler, 2007). Having a specific system for representing objects from an early age can be beneficial because it allows for the incorporation of prior knowledge and expectations about objects and their physical regularities, such as the idea that objects usually maintain their shape and size as they move (*rigidity* principle; Spelke, 1990) and continue to exist and retain their properties even when occluded (*object persistence* principle; Baillargeon, 1987, 2008). Despite converging evidence for the existence of a core object system in both human infants (e.g., Feigenson & Carey, 2003; Spelke, 2022) and non-human animals (e.g., Chiandetti, Spelke, & Vallortigara, 2015; Hauser & Carey, 2003), it is not clear if a system specifically designed for this purpose is necessary or beneficial if object representations can be learned effectively by a domain general inductive system from only a small amount of data.

[1] Project page: janphilippfranken.github.io/object-perception

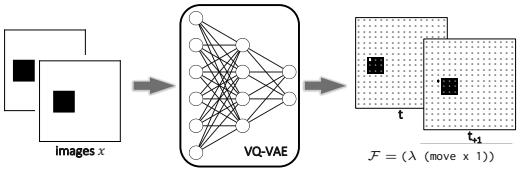

Figure 1: Illustration of inference pipeline. We take short videos (ten frames) as input. These are preprocessed into a sequence of discrete feature maps using vector quantization followed by one-hot encodings on each feature map to obtain Boolean tensors (or "bitmasks"). Bitmasks are then processed by a generic Bayesian concept learning algorithm to induce programs that parsimoniously explain the underlying structure in the discretized data. For example, evaluating the program $\mathcal{F}$ will move ("roll") the upper bitmask (t) by 1 on the $\mathtt{x}$ dimension predicting the bitmask shown below (t_{+1}).

Infant object perception as program induction

We assume that object representations can be used to efficiently compress and discretize perceptual input (e.g., visual). As such, object representations might arise from reasoning about the physical regularities (Spelke, 1990) or "invariants" (Sloman & Lagnado, 2004) in one's environment that facilitate predictions about its future states. Inspired by recent advances in Bayesian program learning (Ellis et al., 2021; Tang & Ellis, 2022; Yang & Piantadosi, 2022) and intuitive physics (Piloto, Weinstein, Battaglia, & Botvinick, 2022), we present an idealized model (Figure 1) that discovers object representations and their physical regularities from short sequences of 2D images (which should generalize to 3D scene projections). Our model can be summarized in four steps: (1) Extract a discrete "codebook" representation c for each image x using a VQ-VAE (Van Den Oord, Vinyals, et al., 2017), a simple tool for efficient image encoding without relying on semantic object assumptions $\rightarrow$ (2) Apply n deterministic one-hot encodings to each discrete feature map c to generate Boolean tensor representations ("bitmasks"), with n representing the number of unique codes in c $\rightarrow$ (3) Use a Bayesian concept learning algorithm to process the resulting bitmasks and generate programs that parsimoniously explain the structure in the data $\rightarrow$ (4) Use discovered programs to improve the representation by searching for structure in residuals or imputing missing data to maximize likelihood. The final two steps are repeated until convergence or until a time-out threshold is reached. To discover programs, our model generates compositions of functions from the primitives listed in Table 1 and computes posterior distributions over programs using Bayes' rule: $P(H \mid D) \propto P(H)P(D \mid H)$. The prior probability of a program $P(H)$ is determined by a probabilistic context-free grammar (PCFG) based on the operations in Table 1. For likelihood $P(D \mid H)$ we assume a standard exponential loss function. We use stochastic (MCMC) sampling as in Goodman, Tenenbaum, Feldman, and Griffiths (2008) to search for programs.

Table 1. Assumed primitive functions

Type	Primitives
Number functions	`(add n)`, `(sub n)`, `(mult n)`, `(div n)`, `(mod n)`, `(neg)`, `(const)`
Set functions	`(union)`, `(intersection)`
Bitmask functions	`(move x n)`, `(move y n)`, `(complement)`, `(const)`

The space of programs consists of all compositions of these functions that respect the input and output types.

Experiments

We evaluate our model's ability to learn object representations and their regularities across eight micro-worlds inspired by experiments from the developmental literature. We use ten images for each probe. To demonstrate our approach, we first examine a baseline probe including simple left-right movement (Fig. 2a). To show that we can also handle natural categories that violate standard object properties, we next consider a "melting" block (i.e., a block that is shrinking vertically; Fig. 2b). We then test the ability of our approach to discover, from sparse input, principles that are often considered as core knowledge, including the widely studied principles of object persistence (Baillargeon, 2008; Piloto et al., 2022; Fig. 2c–d) and rigidity (Spelke, 1990; Kemp & Xu, 2008; Fig. 2e–g) . We additionally include an example of unchangeableness following occlusion (Baillargeon & Carey, 2012; Fig. 2h).

Results

Panels a–b in Fig. 2 show that our model can find programs capturing simple object regularities such as constant left-right movement, which can be expressed as `(λ (move x n))` as well as "melting" which can be expressed as `(λ (intersection (move y (neg n)) (const)))`. Figure 2c shows that this ability still holds for an object that moves behind an occluder. Figure 3a shows the probability of the occluded object for frame 5 in Figure 2c. Consistent with a flattening learning curve at 1000 samples, the model is learning representations at around 1000 samples. The representation of the occluded object was obtained by imputing its representation using a program such as `(λ (move x n))` which will have a maximum likelihood if it keeps representing the object during occlusion. In line with this idea, the example in Figure 2d has a weaker learning curve as the object does not reappear, making it harder to find a physically plausible regularity of the object. Overall, these findings are consistent with increased surprise in infants when objects suddenly disappear or reappear after obstacles as well as their tendency to keep representing objects during occlusion (Baillargeon, 2008).

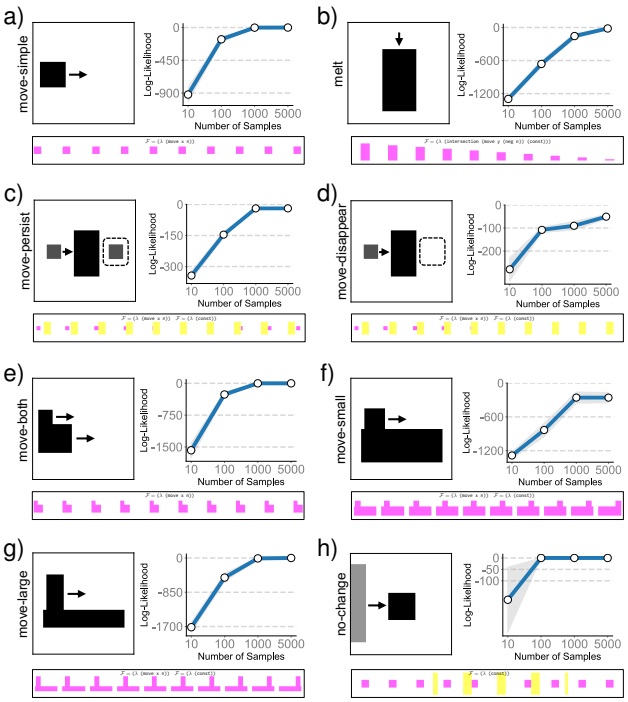

Figure 2: Illustration of tested micro-worlds and learning curves for the model. The x axis corresponds to the number of samples (i.e., the length of the MCMC chain) and the y axis corresponds to the log-likelihood of the final program(s) from a given chain averaged across 100 independent runs. Grey shadings correspond to standard error. Full sequence for each micro-world and target program(s) are shown at the bottom of each panel.

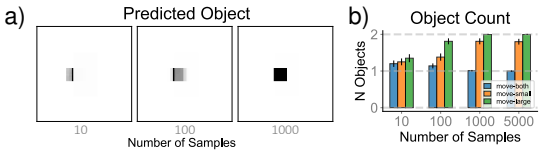

Figure 3: a) Probability of predicted object (greyscale) during occlusion for different numbers of samples. b) Average object counts ($\pm$ SEM) for the three example tests of rigidity.

Results in Figure 2e–g demonstrate our model's ability to interpret ambiguous scenes involving two blocks of different sizes similar to how infants do (Kestenbaum, Termine, & Spelke, 1987; Spelke, von Hofsten, & Kestenbaum, 1989). Specifically, our model provides a single-object interpretation for the example shown in Figure 2e and a two-object interpretation for the examples shown in Figure 2f-g, which require two regularities (both `(λ (move x n))` and `(λ (const))`). Average object counts for different numbers of samples are shown in Figure 3b, and stable performance is achieved around 1000 sample. Our final example (Figure 2h) demonstrates the concept of unchangeableness where an object is occluded by a plank moving across the scene (we do not model the plank's regularity). Following the same imputation approach as in Figure 2c, our model can efficiently learn the objects regularity from a small amount of data.

Discussion

Object representations form a fundamental aspect of human and machine cognition. We proposed that these representations can be learned by domain general learning system that aims to induce symbolic programs to maximize the likelihood of observations. A limitation of the present proof-of-concept results is that we did not jointly train the VQ-VAE and search for programs but instead trained the VQ-VAE prior to search to obtain discrete codebooks. Future work should thus explore joint end-to-end learning of both the VQ-VAE and program search to test our model in more complex scenes (e.g., Piloto et al., 2022; Mao, Yang, Zhang, Goodman, & Wu, 2022).

References

Baillargeon, R. (1987). Object permanence in $3 1/2$-and $4 1/2$-month-old infants. *Developmental psychology*, *23*(5), 655.

Baillargeon, R. (2008). Innate ideas revisited: For a principle of persistence in infants' physical reasoning. *Perspectives on Psychological Science*, *3*(1), 2–13.

Baillargeon, R., & Carey, S. (2012). Core cognition and beyond: The acquisition of physical and numerical knowledge.

Chen, H., Venkatesh, R., Friedman, Y., Wu, J., Tenenbaum, J. B., Yamins, D. L., & Bear, D. M. (2022). Unsupervised segmentation in real-world images via spelke object inference. *arXiv preprint arXiv:2205.08515*.

Chiandetti, C., Spelke, E. S., & Vallortigara, G. (2015). Inexperienced newborn chicks use geometry to spontaneously reorient to an artificial social partner. *Developmental Science*, *18*(6), 972–978.

Ellis, K., Wong, C., Nye, M., Sablé-Meyer, M., Morales, L., Hewitt, L., ... Tenenbaum, J. B. (2021). Dreamcoder: Bootstrapping inductive program synthesis with wake-sleep library learning. In *Proceedings of the 42nd acm sigplan international conference on programming language design and implementation* (pp. 835–850).

Feigenson, L., & Carey, S. (2003). Tracking individuals via object-files: evidence from infants' manual search. *Developmental Science*, *6*(5), 568–584.

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

Hauser, M. D., & Carey, S. (2003). Spontaneous representations of small numbers of objects by rhesus macaques: Examinations of content and format. *Cognitive Psychology*, *47*(4), 367–401.

Kemp, C., & Xu, F. (2008). An ideal observer model of infant object perception. *Advances in neural information processing systems*, *21*.

Kestenbaum, R., Termine, N., & Spelke, E. S. (1987). Perception of objects and object boundaries by 3-month-old infants. *British journal of developmental psychology*, *5*(4), 367–383.

Mao, J., Yang, X., Zhang, X., Goodman, N., & Wu, J. (2022). Clevrer-humans: Describing physical and causal events the human way. In *Thirty-sixth conference on neural information processing systems datasets and benchmarks track*.

Piloto, L. S., Weinstein, A., Battaglia, P., & Botvinick, M. (2022). Intuitive physics learning in a deep-learning model inspired by developmental psychology. *Nature human behaviour*, *6*(9), 1257–1267.

Schölkopf, B., Locatello, F., Bauer, S., Ke, N. R., Kalchbrenner, N., Goyal, A., & Bengio, Y. (2021). Toward causal representation learning. *Proceedings of the IEEE*, *109*(5), 612–634.

Sloman, S., & Lagnado, D. A. (2004). Causal invariance in reasoning and learning. *Psychology of learning and motivation*, *44*, 287–326.

Spelke, E. S. (1990). Principles of object perception. *Cognitive science*, *14*(1), 29–56.

Spelke, E. S. (2022). *What babies know: Core knowledge and composition volume 1* (Vol. 1). Oxford University Press.

Spelke, E. S., & Kinzler, K. D. (2007). Core knowledge. *Developmental science*, *10*(1), 89–96.

Spelke, E. S., von Hofsten, C., & Kestenbaum, R. (1989). Object perception in infancy: Interaction of spatial and kinetic information for object boundaries. *Developmental Psychology*, *25*(2), 185.

Tang, H., & Ellis, K. (2022). From perception to programs: regularize, overparameterize, and amortize. *arXiv preprint arXiv:2206.05922*.

Van Den Oord, A., Vinyals, O., et al. (2017). Neural discrete representation learning. *Advances in neural information processing systems*, *30*.

Xu, F., & Carey, S. (1996). Infants' metaphysics: The case of numerical identity. *Cognitive psychology*, *30*(2), 111–153.

Yang, Y., & Piantadosi, S. T. (2022). One model for the learning of language. *Proceedings of the National Academy of Sciences*, *119*(5), e2021865119.