Title
Reverse engineering the origins of visual intelligence

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 42(0)

Authors
Wood, Justin N.
Lee, Donsuk
Wood, Brian W.
et al.

Publication Date
2020

Peer reviewed

Reverse engineering the origins of visual intelligence

Justin N. Wood[1-3] (woodjn@indiana.edu)**, Donsuk Lee**[1] (donslee@indiana.edu)**,**
Brian W. Wood[1] (bwwood@indiana.edu)**, Samantha M. W. Wood**[1] (sw113@indiana.edu)

[1]Department of Informatics, Bloomington, IN 47408 USA
[2]Department of Cognitive Science, Bloomington, IN 47408 USA
[3]Center for the Integrative Study of Animal Behavior, Bloomington, IN 47408 USA

Keywords: visual development; newborn; controlled rearing; innate; object recognition; machine learning; ANN models

Introduction

In recent years, researchers have made great strides developing a mechanistic understanding of object recognition in mature brains. Despite this progress, fundamental questions remain about the origins and development of object recognition. To what extent is the 'initial state' of object recognition innately constrained? What are the learning algorithms that transform the initial, naïve state into a mature state? Here, we describe a new experimental approach for studying the origins and development of object recognition, by performing parallel controlled-rearing experiments on newborn chicks and autonomous artificial agents. This approach can be used to isolate the core computational components that underlie visual intelligence.

The Learning Problem

What are the core learning mechanisms in newborn brains? What role does experience play in calibrating those mechanisms over time? Despite significant interest in these questions, the field lacks a mechanistic understanding of how visual intelligence develops in newborn brains.

In computational neuroscience, the vast majority of studies have compared artificial neural network (ANN) models to *mature* visual systems (e.g., Schrimpf et al., 2018). ANN models are typically trained with supervised learning, in which the model learns from millions of labeled training images. However, animals learn largely through unsupervised learning. Human infants receive labeled object input only when they begin understanding language, and nonhuman animals receive no labeled training input during development. Building accurate models of the visual system therefore requires building ANN models that *learn* like newborn brains, using unsupervised learning algorithms.

There are three challenges in building ANN models that learn like newborn brains. First, newborn subjects are hard to study. Most methods for studying cognition in newborn subjects are low powered and produce noisy measurements. This makes it challenging to obtain the precise benchmarks needed to build accurate ANN models. Second, it is not possible to control the environment in which most newborn subjects are raised. Thus, we do not know which visual experiences were used to 'train' their visual system during development. Since the outputs of ANN models change radically as a function of training input, accurate comparisons of ANN models and newborn brains require training the models and brains with the same visual experiences. Third, animals have bodies and choose their own input during development through active exploration. To directly compare ANN models and newborn brains, we must *embody* ANN models and allow the models to choose their own input during learning.

To overcome these three challenges, we developed two new tools for reverse engineering the origins of object recognition. These tools allow us to obtain precise benchmarks from newborn animals (§ 1) and then use those benchmarks to directly compare the learning abilities of embodied ANN models and newborn brains (§ 2).

1. Controlled Rearing Studies of Newborn Chicks

Reverse engineering the origins of object recognition requires precise benchmarks showing how specific visual inputs shape the development of object recognition. To obtain these benchmarks, we use newborn chicks as a model system. Unlike most animals, chicks are uniquely suited for studying the earliest stages of visual development (Wood & Wood, 2015). Chicks are mobile on the first day of life, require no parental care, and can be raised in strictly controlled environments immediately after hatching. Thus, it is possible to study how specific experiences shape object recognition. Since cortical mechanisms are largely conserved across birds and mammals (e.g., Karten, 2013), studies of chicks can also reveal general insights into vertebrate visual development.

We developed a high-powered controlled-rearing method for studying newborn object recognition, using automated image-based tracking software and virtual stimuli. We raise newborn chicks in strictly controlled virtual worlds and record their behavior 24/7 as they learn to recognize objects. Computers perform all stimuli presentation and behavioral coding, so we can monitor the chicks' behavior continuously across the first 3-4 weeks of life. As a result, we can obtain precise measurements of performance, chart how object recognition changes over time, and examine individual differences across newborn subjects. Fueled by video game engines, we can also raise chicks in interactive virtual reality worlds with realistic 3D objects and scenes.

We have made discoveries in two areas. First, invariant object recognition can emerge rapidly in newborn brains. For instance, when newborn chicks are reared in a world with a single object, they can recognize that object across novel viewing situations, including changes in viewpoint, background, and motion speed (Wood, 2013; 2015; Wood & Wood, 2015; 2016). Newborn chicks can also perform "one-shot learning." When chicks are reared with a single view of a single object, they can recognize that object across novel viewpoints (Wood & Wood, 2020). Likewise, when chicks are reared with a single object on a single background, they can recognize the object across novel backgrounds.

These results compliment a growing body of work showing that biological brains can learn new concepts from just one or a handful of examples (e.g., Lake, Salakhutdinov, & Tenenbaum, 2015). The discovery that newborn chicks can perform one-shot learning suggests that this ability does not require extensive post-natal experience in order to develop. One-shot learning appears to scaffold object recognition during the earliest stages of learning.

The second area of discovery is that newborn brains need a specific type of training data in order to develop invariant object recognition; namely, experience with a naturalistic visual environment (Wood, 2016; Wood & Wood, 2016; 2018; Wood, Prasad, Goldman, & Wood, 2016). To learn correctly, newborn chicks need input of object views that change slowly and smoothly over time, adhering to the spatiotemporal properties of objects in the real world. Without slow and smooth visual input, chicks build 'incorrect' object representations that fail to generalize across new viewing situations. Thus, newborn brains learn to see by leveraging the slow and smooth input from natural visual environments—a key prediction of unsupervised temporal learning models in computational neuroscience.

2. Controlled Rearing Studies of Artificial Chicks

The methods and results described above provide the foundations for reverse engineering the origins of object recognition. By using high-powered methods to study the origins of vision, we can collect precise data showing how newborn brains transform sensory inputs into behavioral outputs. These input-output patterns can then serve as benchmarks for building end-to-end (pixels-to-actions) artificial agents that learn how to see like newborn animals.

To this end, our lab developed a machine learning platform for linking biological intelligence to artificial intelligence. This platform allows us to raise newborn animals and artificial agents in the same environments, and test their visual recognition behavior with the same tasks.

Our project involves three steps: First, we build artificial brains (deep neural networks) with visual systems that learn through biologically-plausible mechanisms (e.g., predictive coding, episodic memory, curiosity-driven learning). These brains are embodied in virtual chick bodies. Second, we raise the artificial chicks in the same virtual reality environments as real chicks, and test their behavior with the same tasks. Thus, we directly compare the learning abilities of biological brains and artificial brains that have received the same training data. Third, by plugging different artificial brains into the artificial chicks, we can isolate the core learning mechanisms that underlie visual intelligence.

Using this approach, we have recently identified a class of ANN models that develops some of the same visual abilities as newborn brains. When artificial brains contain deep neural networks that learn through curiosity-driven learning, and the artificial chicks are raised in realistic environments, the artificial chicks spontaneously develop ego-motion, object parsing, and object recognition. Moreover, the artificial chicks spontaneously learn to detect and imprint to animate agents. We argue that a relatively simple set of core computational components is sufficient to produce foundational visual behaviors.

Acknowledgments

Funded by NSF CAREER Grant BCS-1351892 and a James S. McDonnell Foundation Understanding Human Cognition Scholar Award.

References

Karten, H. J. (2013). Neocortical evolution: neuronal circuits arise independently of lamination. *Current Biology, 23*(1), R12–5.

Lake, B. M., Salakhutdinov, R., & Tenenbaum, J. B. (2015). Human-level concept learning through probabilistic program induction. *Science*, *350*(6266), 1332–1338.

Schrimpf, M. et al. (2018). Brain-Score: Which artificial neural network for object recognition is most brain-like? *bioRxiv*, *6*(11), 27755–9.

Wood, J. N. (2013). Newborn chickens generate invariant object representations at the onset of visual object experience. *PNAS*, *110*(34), 14000–14005.

Wood, J. N. (2015). Characterizing the information content of a newly hatched chick's first visual object representation. *Developmental Science*, *18*(2), 194–205.

Wood, J. N. (2016). A smoothness constraint on the development of object recognition. *Cognition*, *153*, 140–145.

Wood, J. N., Prasad, A., Goldman, J. G., & Wood, S. M. W. (2016). Enhanced learning of natural visual sequences in newborn chicks. *Animal Cognition*, *19*(4), 835–845.

Wood, S. M. W., & Wood, J. N. (2015). A chicken model for studying the emergence of invariant object recognition. *Frontiers in Neural Circuits*, *9*(89), 7.

Wood, J. N., & Wood, S. M. W. (2016). The development of newborn object recognition in fast and slow visual worlds. *Proceedings of the Royal Society: Biological Sciences,* 283, 20160166.

Wood, J. N., & Wood, S. M. W. (2018). The development of invariant object recognition requires visual experience with temporally smooth objects. *Cognitive Science*, *39*(17), 2885–16.

Wood, J. N. & Wood, S. M. W. (2020). One-shot learning of view-invariant object representations in newborn chicks. *Cognition, 199,* 104192.