

What Do Babies Have That Computers Don't?

March 29, 2019



Technology breakthroughs have enabled machines to recognize and respond to our voices, identify our faces, and even translate text written in another language. Despite all the research funding and venture capital that have been poured into these advances, artificial intelligence is still easily stymied in novel situations and remains limited in its grasp of natural language.

APS William James Fellow Linda B. Smith believes machine learning could transcend some of these shortcomings by mimicking the learning processes of infants and toddlers.

So what does a child have that a computer lacks? In her 2018 award address at the 30th APS Annual Convention in San Francisco, Smith explained how the sophistication of human visual learning enables babies to grasp the names and categories of objects in ways that are thus far unmatched in the world of artificial intelligence.

To illustrate, she used the example of a 2-year-old child seeing a John Deere tractor operating in a field for the first time.

“If the child watches that tractor work and is told, ‘It’s a tractor, it’s a tractor, it’s a tractor,’ it’s highly

likely that from that point forward that 2-year-old will recognize all varieties of tractors — Massey Fergusons, antique tractors, ride-on mowers — but will not consider a tank or a backhoe to be a tractor,” she said.

In developmental psychology, this phenomenon is known as the shape bias — the tendency to generalize information about objects by their shapes rather than by their colors, sizes, or other physical attributes. In the machine-learning literature, that same phenomenon is known as one-shot category learning — the ability to take information about a single instance of a category and extrapolate it to the whole category.

Children are not born with this one-shot ability; they learn it within their first 30 months of life, developmental studies have demonstrated. Smith, principal investigator at Indiana University Bloomington’s Cognitive Development Lab, is among researchers who have studied training exercises that can encourage the shape bias to emerge 6 to 10 months earlier than normally expected.

The exploration of early language development, Smith explains, centers on the two parts of the learning process: the training data and the mechanisms that do the learning.

“The data for learning, the experience on which visual category learning occurs in infants, is fundamentally different from the experiences that are used in machine learning to train computer vision — and from the experiences that are used in experimental psychology studies of learning,” Smith said.

Those differences, she said, may help explain why the human visual system is so sophisticated, and why babies “can learn object names in one shot.”

From the Eyes of Babes

Smith has employed a variety of methods to study linguistic development and object learning. But one of her best-known approaches is the use of head-mounted video cameras, eye trackers (typically embedded in a hat or headband), and motion sensors to view a child’s visual world from his or her own point of view. The bulk of this work is being conducted through the Home-View Project, an initiative that Smith and her Indiana University colleague Chen Yu developed with support from the National Science Foundation.

The Home-View Project has collected nearly 500 hours of video involving more than 100 babies, who range in age from 3 weeks to 24 months, as they go about their daily lives at home. The data collected to date show that babies learn a massive amount of information based on just a few faces, objects, and categories, with that learning changing at different points in time. They generate their own data for learning based on how they move and position their bodies. In the first few months of life, when they possess little control of the head and body, they’re mainly seeing the faces of their parents and caregivers. But as they approach their first birthday, they focus more of their attention on hands and objects.

In all these domains, they learn a lot about a few specific things — their mother’s face, their sippy cup, the family dog. At the same time, they’re learning “a very little about lots and lots of other stuff,” Smith said.

Her recent experiments indicate that babies learn the names of objects based on the prevalence of those objects in their visual world. In one study, the results of which were published in 2016, Smith and her team of researchers examined videos that showed the visual field of babies ages 8 to 10 months. The children wore the cameras at home for an average of 4.4 hours per day.

The researchers focused their observation on hours of mealtime scenes, Smith explained.

“We counted as mealtime any event that had food or dishes in it,” she said. “If a dog was eating, that’s a mealtime. Cheerios on the living room floor — mealtime.”

Although most scenes were cluttered, a few objects (e.g., a chair, a spoon, a bottle) were the most frequent items in the child’s visual experience. And with this approach, the researchers could identify when the children learned names for object categories and individual objects. Results showed that the first nouns the children learned centered on the objects they saw most consistently.

“This suggests to us that visual pervasiveness itself — day in, day out, hours upon hours, from many different viewpoints — may be critical to visual learning about objects, to building segmentation, to finding things in clutter, to building strong visual memories so that you can eventually get a word attached to them,” Smith says.

The World in Their Hands

Her experiments are also examining how babies’ visual experiences with contrast and light change over time, and how the eventual experience of engaging objects with their hands factors into their object-name learning. By the time infants reach their first birthday, they’re beginning to control what they see by handling objects, not just looking at them.

“By holding [an object], looking at it, and mom naming it for them, toddlers create clean images of single objects that dominate in the scene,” Smith said. “When parents name objects at those clean moments, the child is much more likely to learn the object name.”

Smith’s research is now examining the roles that culture and socioeconomics play in these processes. A parallel Home-View Project in India, for example, is working with children in a remote village that lacks electricity.

The research has left Smith confident that machines may indeed become one-shot category learners if they’re simply fed infant head-camera images. Understanding the roles of environment and visual experiences also could lead to new interventions for children with conditions such as autism spectrum disorder, which are associated with language and visual learning deficits.

“Could we alter the developmental outcomes for children,” she asked, “by providing differently structured day-in and day-out data for learning, giving them a workaround if something’s amiss?”