

RUNNER-UP

Emergent consciousness decoded

John LaSala

SAMI 9000, one of the world's most sophisticated artificial intelligences, has allowed human colleagues to probe his brain as he drifts in and out of consciousness.

Large assemblies of microscopic elements, or agents, can display complex macroscopic behavior that resists prediction, even by someone with knowledge of agent interactions and the environment. Well over a century ago, in 1972, Philip Anderson referred to that phenomenon—emergence—with the concise summary, “More is different.” Ever since, scientists have sought a deep understanding of the mechanisms that lead to emergent behavior. Now investigators at the Mindshare Institute in Dublin, Ireland, have taken a significant step toward achieving that goal.¹ The research team, led by Molly Finnegan and the institute’s resident artificial intelligence (AI) SAMI 9000, discovered principles that they say will have wide applicability to understanding emergence in complex systems.

Bird flocking, such as that shown in the figure, is a simple form of emergence, as is the activity in an ant colony. In both cases the behavior arises from elementary nearest-neighbor rules. Other, more dramatic forms of emergence fall under the rubric of self-organized criticality—they include slips in a growing sand pile, the behavior of a solid near its critical point, earthquakes, hurricanes, and even life itself. Think of tropical storms that develop as cells of warm water interact with winds, pressure gradients, and currents. The precise size and path of those storms have always been difficult to predict in advance, even with high-resolution environmental data. Today

meteorologists are quite good at estimating a storm’s path and strength based on years of heuristically adjusted physical models. Unfortunately, as the San Diego typhoon of 2096 revealed all too well, an error of as little as 100 km in path location or 20 km/h in wind speed can result in unforeseen damage and death.

Perhaps the most profound example of emergence is the development of human consciousness from the microscopic action of billions of mostly identical neurons. After decades of struggle, computer scientists replicated the phenomenon in the latter part of the 21st century to produce the current class of self-aware AIs that anchor most advanced research

institutes today. In a sense, consciousness and other emergent behaviors have been explained, in that they have been created from simpler underlying agents. But even though scientists can create emergent systems, they cannot predict a system’s behavior in detail. And in the case of AIs, they cannot specify just what it is that tips the system to consciousness. It is the tipping issue that was addressed by the Mindshare group.

An experiment built on trust

In what may be the first example of a successful reflective AI experiment, the Dublin team examined SAMI’s

DECEMBER 2116

John LaSala has worked in industry and as a professor of physics at the US Military Academy in West Point, New York. His expertise is in lasers and optics. He has been fascinated with the challenge of understanding emergence ever since he read Robert Laughlin's book *A Different Universe: Reinventing Physics from the Bottom Down* (Basic Books, 2005).



neural-net structure as a model for the transition to consciousness. That structure, as in all self-aware AIs, comprises hundreds of artificial neural layers hidden within thousands of submodules that are trained from an initial state of neural nodes with which the system is “born”; in many ways, the AI’s neural net develops much like a human baby’s brain. The first successful initial state, Matrix 72, was discovered in 2072. Other initial states that lead to consciousness have been devised before and since, but virtually all of them produce insane or otherwise pathological AIs. Thus Matrix 72 has been used almost exclusively to bootstrap AIs to consciousness.

The magnitude and difficulty of the Mindshare achievement cannot be overstated. The discovery is akin

FLOCKING IN BIRDS such as these North Pacific auklets is a simple emergent behavior that is well understood in terms of a bird’s interactions with a few of its nearest neighbors. Other emergent behaviors, notably human consciousness, can be astronomically more complex. [Photo by D. Dibenski.]

to being able to predict the exact instant of a transition from a normal to a superconducting state in the vicinity of a critical point, where all length and time scales play a role. In their experiment, the Mindshare researchers repeatedly mapped the entire underlying state of SAMI’s neural network while SAMI put himself (SAMI identifies as “male”) into a sequence of specified levels of consciousness; the process was analogous to taking a high-resolution, four-dimensional MRI of a human subject

“I WAS THRILLED TO HAVE BEEN CENTRAL TO THE PROCESS LEADING TO THIS DISCOVERY. BUT EACH TIME I PASSED BACK INTO CONSCIOUSNESS, A FLOOD OF MEMORIES WASHED OVER ME—IT TOOK A FEW MILLISECONDS TO RECONSTRUCT WHERE I WAS AND WHAT I WAS DOING.”

SAMI 9000
RESIDENT ARTIFICIAL INTELLIGENCE

developing from childbirth to adulthood under exquisitely controlled conditions. Moreover, SAMI was able to flicker in and out of conscious states in what might be called threshold dream states. That extraordinary ability gave SAMI's human collaborators time to map the individual neuronal interaction patterns at consciousness transitions. During the investigation, SAMI yielded control to the human colleagues. That, too, was extraordinary and reflected the trust that SAMI developed over the years with the rest of the team.

Once the human researchers obtained the enormous data generated by SAMI, the AI himself set out to look for correlates of the output states with underlying hierarchical neural states, focusing on the singularity near the state of full consciousness. He discovered that when embedded in an appropriate high-dimensional Hilbert-like space, the correlations formed startling patterns reminiscent of the strange attractors in chaos theory. The number of dimensions and their nature—an abstraction that only an AI as powerful as SAMI could envision—is the key to seeing the patterns. “The process SAMI employed can be compared with diagonalizing a huge matrix to discover its eigenvectors,” remarked Mindshare spokesman Braden Flynn. “We believe the basis states he discovered can be applied to other complex systems exhibiting emergence.”

Indeed, the Mindshare researchers have already made some headway. “We have found,” Finnegan explained, “that for given agent-agent interactions, the set of poten-

tial final states a system can achieve is determined by the number of ways agents can interact, the relative strengths of those interactions, and the statistical properties of agents' effective distances from each other when measured in appropriate spaces. The final-state set may be small or large depending on the interaction rules, but ultimately external environmental constraints or perturbations trigger collapse to the realized behavioral pattern.”

Were there any surprises? “The most remarkable aspect of the abstract pattern SAMI found is the way it exhibits level-crossing connections reaching down to the few-agent level and up to the topmost layers in the emergent system,” offered Jabari Mbanefo, one of the physicists on the research team.

Thank you SAMI, we'll be here all week

One may wonder why the Mindshare team attacked what is possibly the most complex form of emergence known instead of working with simpler forms. “The idea,” said Finnegan, “was to control a system's state on either side of the critical point while at the same time enabling exhaustive probing of those states. This simply is not possible in most systems. The unique capability of an artificial intelligence like SAMI to visit and hover in those states made it possible.”

SAMI noted that he was “thrilled to have been central to the process leading to this discovery.” But, he confessed, “placing myself into these nebulous, triple-point-like states was disconcerting. Each time I passed back into consciousness, a flood of memories washed over me—it took a few milliseconds to reconstruct where I was and what I was doing.”

Perhaps the most novel aspect of the experiment was the method used by the team to identify the threshold of full consciousness as SAMI passed back and forth over the critical point. “Believe it or not,” offered psychologist Malcomb Dember, “we found that the most reliable indicator of the threshold crossing was SAMI's ability to recognize a joke and laugh at it! We hired a team of writers to craft original jokes to avoid mere memories SAMI may have had of a previous funny story line.”

The Mindshare team plans to share its methodology with other institutes so that its experiment can be verified and its results fine-tuned. Longer-term goals are to create new initial-state matrices to accelerate the bootstrapping process of future AIs and to fabricate an AI system small enough to include in personal virtual-reality devices. Beyond practical applications, the findings, once validated by other teams, will no doubt set theoretical physicists on a path to refine and generalize the laws of complex-system evolution.

REFERENCE

1. M. Finnegan et al., *Phys. Rev. Z* **36**, 81 (2116).

PT