Bionoetics, vol. 1, no. 1, 2014

INFORMATION, PHYSICS AND THE REPRESENTING MIND

Kathryn Blackmond Laskey

ABSTRACT: A primary function of mind is to form and manipulate representations to identify and choose survival-enhancing behaviors. Representations are themselves physical systems that can be manipulated to reason about, predict, or plan actions involving the objects they designate. The field of knowledge representation and reasoning (KRR) turns representation upon itself to study how representations are formed and used by biological and computer systems. Some of the most versatile and successful KRR methods have been imported from computational physics. Features of a problem are mapped onto dimensions of an imaginary physical system in which solution quality is inversely related to energy. Simulating the fictitious physical system on a digital computer yields a low-energy, and hence high-quality, solution to the original problem. This paper suggests a rethinking of the traditional metaphor of cognition as execution of algorithms on a digital computer. It may be both more fruitful and more accurate to conceive of representation as mapping problem features to an energy surface, learning as identifying representations that map good solutions to low free energy, and problem solving as efficient search for low free energy states. This conception of cognition is in natural accord with Stapp's theory of efficacious conscious choice.

KEYWORDS: Knowledge representation and reasoning; Markov Chain Monte Carlo; Physical Symbol System; Quantum Zeno Effect

INTRODUCTION

Representing the world is a primary function of the mind. The ability to form, manipulate and use representations is not unique to humans: honeybees use dance to represent and communicate the direction and distance to food sources [1]; crows, dogs and other animals are capable of making and using tools [2], [3], [4]; several studies indicate that some species of birds and mammals are capable of planning [5], [6], [7]. Human use of representation has fundamentally altered the world around us. The agricultural revolution dramatically increased food production, generating an exploding human population and the rise of cities [8], [9], [10]. The industrial revolution reorganized production of material goods, transforming human society and radically altering the global environment [11]. Now, as the information revolution unfolds around us [12], humanity is applying our representation capability to representation itself. Although the full impact is not yet manifest, the information revolution is clearly giving rise to disruptive change on a worldwide scale. Each of these major worldwide revolutions has grown out of our ability to construct representations, manipulate those representations to form plans and understand their effects, execute the plans, and build improved representations using feedback on successes, failures and unintended side effects.

Underlying the information revolution is scientific study of the phenomenon of representation itself. Shannon's influential theory of information has found wide application [13]. Database technology, concerned with representing, storing and accessing information in computers [14], [15], [16], is a critical element of the infrastructure of today's business enterprise. Artificial intelligence employs computational knowledge representations to allow computers to exhibit intelligent behavior on tasks once thought to require humans [17]. Cross-fertilization between artificial intelligence and cognitive science has resulted in computational theories of human cognition and, conversely, artificial intelligence formalisms inspired by empirical research on human problem solving [18], [19], [20].

Knowledge representation and reasoning methods have grown more sophisticated as the problems have grown more challenging. Some of the most successful information processing methods originated in computational physics. The key insight of these physics-based methods is to represent the space of possible solutions as a fictitious multi-dimensional physical space in which good solutions have low energy. The problem of finding a good solution to a problem is thus transformed into the problem of finding a low-energy state in a physical state space. This enables the application of techniques from computational physics to solve information problems.

The successful analogy with computational physics may reflect something fundamental about how organisms form and manipulate representations and perform goal-directed action. This paper suggests a rethinking of the traditional metaphor of cognition as execution of algorithms on a digital computer. It may be both more fruitful and more accurate to conceive of representation as mapping problem features to an energy surface, learning as identifying representations that map good solutions to low free energy, and problem solving as efficient search for low free energy states. This conception of cognition is in natural accord with Stapp's [21] theory of efficacious conscious choice.

PHYSICAL SYMBOL SYSTEMS

Newell and Simon [18] pioneered the now-common practice of developing computational theories of intelligent behavior, implementing the theories as computer programs, and evaluating the theories by comparing with human problem solving behavior. They stressed that cognition is performed by the physical brain and nervous system of an embodied agent situated in a physical environment. They offered the *physical symbol system hypothesis* as a scientific, empirically testable hypothesis about the nature of intelligence. The hypothesis states: "A physical symbol system has the necessary and sufficient means for intelligent action." As defined by Newell and Simon, a physical symbol system is:

> ... a set of entities, called symbols, which are physical patterns that can occur as components of another type of entity called an expression (or symbol structure). [Symbols in a structure] are related in some physical way… A physical symbol system … produces through time an evolving collection of symbol structures.

They go on to say that symbol structures can *designate* objects external to the system, and that a physical symbol system can *interpret* a designated process (i.e., invoke and execute it). Thus, a physical symbol system can reason about the consequences of available actions, make a choice, and then either execute the chosen action itself or instruct an external system to execute it.

Newell and Simon clearly intended for digital computers to qualify as physical symbol systems. Smartphones, consumer devices with embedded computers, and the World Wide Web also qualify. The practical success of these physical symbol systems is unquestionable. After fewer than three decades of existence, digital cellular technology and the World Wide Web already pervade every aspect of our lives. Computers recommend services and merchandise we did not know we wanted, help us to navigate to our destinations, check our spelling and grammar, translate articles from foreign languages, and even win at *Jeopardy*. Most work in artificial intelligence takes for granted the *weak AI hypothesis*, that computers can be designed with cognitive abilities matching or surpassing that of humans. In fact, in some domains once thought the exclusive purview of humans, computers have already surpassed our abilities. These successes aside, few argue that today's computers actually possess minds. The *strong AI hypothesis*, that a computer executing a program could have a mind in the same sense as a human does, is regarded by most researchers as irrelevant to the goals of artificial intelligence [17].

PHYSICS-INSPIRED METHODS FOR KNOWLEDGE REPRESENTATION AND REASONING

Intelligent systems form representations and manipulate them to find good interpretations of situations and to select courses of action that they predict will lead to desirable outcomes. As researchers have struggled to build artificially intelligent systems, it has become increasingly clear just how difficult a problem this is. A place to which researchers have turned to find good general-purpose problem solving methods is computational physics.

One of the most versatile and widely applicable physics-inspired techniques has become known as Markov Chain Monte Carlo (MCMC). Originally developed to solve problems in statistical physics, MCMC has become a popular general-purpose approximation approach applied to a broad variety of intractable optimization and statistical estimation problems [22]. The basic idea is to map problem features onto dimensions of a fictitious physical system. An energy function is defined such that "good" solutions have low energy and "bad" solutions have high energy. An MCMC sampler makes local moves about the energy surface, its sampling distribution constructed to form a Markov chain with the Boltzmann distribution as its stationary distribution. After a sufficiently long time, therefore, the sampler will have high probability of being found at solutions with low free energy, i.e., "good" solutions.

An early MCMC application outside of physics was an artificial neural network model called the Boltzmann machine [23]. In the Boltzmann machine, as in other physically inspired neural network models, each point on the energy surface corresponds to a different physical state of the artificial brain, which in turn corresponds to the brain's representation of the world. Thus, the MCMC model seeks a low-energy physical brain state, which in turn provides a "good" representation of the world.

Over the past few decades, MCMC has successfully addressed challenging problems in a wide variety of domains such as computer vision, protein sequencing, recommender systems, machine learning, astronomy, robot navigation, and many others. It is fair to say that MCMC has changed the face of applied computer science, enabling solution of many previously intractable problems [24]. The typical MCMC samplers map problem features and solution quality measures to a fictitious energy surface, encode the representation as a data structure on a digital computer, and run the sampler as a computer program, using pseudo-random numbers. Thus, an MCMC algorithm is a digital simulation of a fictitious physical system to which a digital representation of the original problem has been mapped.

THE ENTRY OF MIND INTO PHYSICS

Prior to the twentieth century, mind was regarded as outside the province of science. The causal closure of classical physics seemed to leave mind as a bystander. Thoughts appeared to be byproducts of physical events in the brain with no causal role in how the world unfolds.

The view of mind as outside of science has changed on two fronts: cognitive scientists have developed scientific theories of the mind, and quantum theory forced the entry mind into physical theory. The evolution of a quantum system depends on whether the system is observed, when the observation occurs, and what is observed. While many attempts have been to remove observer dependence from quantum theory, none has achieved broad acceptance. The pragmatic Copenhagen interpretation, which views quantum theory as a construct for organizing the knowledge of observers, has remained the orthodox view. As Heisenberg [25] put it, "The conception of objective reality of the elementary particles has thus evaporated… into the transparent clarity of a mathematics that represents no longer the behavior of particles but rather our knowledge of this behavior."

According to John von Neumann's [26] mathematical formalization of quantum theory, the state of a quantum system is described as a mathematical object called the quantum state, which evolves over time in two different ways. An isolated quantum system exhibits continuous deterministic evolution according to the Shrödinger equation. The second kind of evolution is a discontinuous change called state reduction, or more colorfully, collapse. In both the Copenhagen and von Neumann interpretations, a reduction occurs when an observer interacts with a quantum system, "amplifying" a particular feature of the microscopic quantum world into a macroscopic observation. Quantum theory provides highly accurate predictions of deterministic Shrödinger evolution and highly accurate probabilities for the stochastic outcomes of reductions. However, quantum theory has nothing to say about the time at which a reduction will occur or what, among the allowable options, the set of possible outcomes will be. The founders of quantum theory assigned the choice of time and outcome possibilities to the observer's free choice.

Thus, quantum theory contains an explanatory gap – the choice of time and possible outcomes of reduction. As a pragmatic matter, the founders of quantum theory assigned this gap to the observer's free choice. Stapp [21] goes beyond the pragmatic view to develop a scientific theory of how this free choice operates. Stapp postulates that a conscious agent's intentional choices manifest physically as a choice of time and possible outcomes of a reduction applied to the agent's physical brain. Nature responds to this choice by selecting an outcome stochastically according to the probability rule given by quantum theory. The resulting actual outcome gives rise to an experiential "feel" on the part of the agent.

Stapp's theory breaks the causal closure of classical physics and provides mind with an efficacious role in the evolution of a conscious agent's brain and thereby its body. Developing this idea into a full-fledged scientific theory requires meeting some demanding constraints. Specifically, it must be verified that the choices assigned to the agent by the theory can result in the kinds of physical effects empirically associated with conscious choice. Stapp hypothesizes that conscious choices operate via the quantum Zeno effect (QZE), whereby a sufficiently rapid sequence of state vector reductions can hold a quantum system in place, effectively stopping its evolution. Alternatively, sufficiently rapid application of a different sequence of reductions can be used to drive the system to a desired state, in what has been called the quantum anti-Zeno effect (QAZE). Recent work by Stapp [27] has been directed to establishing that QZE or QAZE can, in settings consistent with warm, wet brains, give rise to macroscopically distinguishable effects that could plausibly lead to observable behavior change.

CONCLUSION

We have seen that many computer and cognitive scientists view physical embodiment as essential to mind; that general-purpose algorithms imported from computational physics have enabled computers to perform in ways some label intelligent; that mind enters in a fundamental way into the orthodox interpretation of our most fundamental theory of physics; and that an explanatory gap in quantum theory provides an opening for a physically well-founded theory of efficacious conscious choice. Tying these threads together suggest that we may be witnessing the beginnings of a unified science of cognition that embraces both the physical and mental aspects of how cognitive agents form, manipulate, and adapt their representations.

Such a unified science would have its basis in new kind of physical symbol system. The model of an algorithm executing on a digital computer would be replaced by a model of an agent using its representation of the world to make goal-directed choices. Such an agent's cognitive apparatus would have a physical aspect and a corresponding mental aspect, in which problem features in the mental representation correspond to dimensions of the physical state, and "good" representations for the problem context map to low-energy physical states. The mental aspects of the theory would be framed in terms of goals, rewards, information flows, sensory inputs, and allowable actions. The physical aspects of the theory would be framed in terms of physical dynamics, free energy minimization, and quantum measurement.

Exploring this idea further, we turn to a MCMC method called Hamiltonian MCMC [28], which was originally developed for estimating weights in a neural network model. In Hamiltonian MCMC, the target energy surface is treated as potential energy, and is augmented with an auxiliary quadratic kinetic energy term. The sampler alternates between a regime of Hamiltonian dynamics based on the total (potential plus kinetic) energy, and a regime of Metropolis-Hastings sampling based on the potential energy. Hamiltonian MCMC is able to take larger steps and thus converge to the target distribution more rapidly than a traditional Metropolis-Hastings sampler. The alternating deterministic and stochastic regimes of Hamiltonian MCMC are reminiscent of the alternating deterministic and stochastic regimes of von Neumann quantum theory.

A quantum variant of Hamiltonian MCMC could form a model of the physical brain and its representation of the world. While virtually all MCMC literature to date has focused on classical samplers simulated on digital computers, quantum implementations of MCMC have been suggested. If Stapp's theory of efficacious conscious choice is correct, then it is not out of the questions that an appropriately constructed physical implementation of a quantum Hamiltonian MCMC would become an instance of Strong AI.

ACKNOWLEDGEMENTS

The author acknowledges the participants of the 2014 Foundations of Mind Conference for stimulating discussions.

SEOR Department,
George Mason University,
Fairfax, VA, 20030,
USA
klaskey@gmu.edu

REFERENCES

[1] J. R. Riley, U. Greggers, A. D. Smith, D. R. Reynolds, and R. Menzel, "The flight paths of honeybees recruited by the waggle dance," *Nature*, vol. 435, no. 7039, pp. 205–207, May 2005.

[2] "Tool Use by a Dingo and a Dog." [Online]. Available: http://www.psychologytoday.com/blog/animal-emotions/201112/tool-use-dingo-and-dog. [Accessed: 13-Apr-2014].

[3] J. H. Wimpenny, A. A. S. Weir, L. Clayton, C. Rutz, and A. Kacelnik, "Cognitive processes associated with sequential tool use in New Caledonian crows," *PloS One*, vol. 4, no. 8, p. e6471, 2009.

[4] R. W. Shumaker, K. R. Walkup, B. B. Beck, and G. M. Burghardt, *Animal Tool Behavior: The Use and Manufacture of Tools by Animals*, Revised and updated edition edition. Baltimore: Johns Hopkins University Press, 2011.

[5] Nicola S. Clayton, Timothy S. Bussey, and A. Dickinson, "Can animals recall the past and plan for the future?," *Nat. Rev. Neurosci.*, vol. 4, pp. 685–691, 2003.

[6] N. S. Clayton, J. M. Dally, and N. J. Emery, "Social cognition by food-caching corvids. The western scrub-jay as a natural psychologist," *Philos. Trans. R. Soc. B Biol. Sci.*, vol. 362, no. 1480, pp. 507–522, Apr. 2007.

[7] N. J. Mulcahy and J. Call, "The performance of bonobos (Pan paniscus), chimpanzees (Pan troglodytes), and orangutans (Pongo pygmaeus) in two versions of an object-choice task," *J. Comp. Psychol.*, vol. 123, no. 3, pp. 304–309, 2009.

[8] G. Barker, *The Agricultural Revolution in Prehistory: Why did Foragers become Farmers?* Oxford: Oxford University Press, USA, 2009.

[9] C. J. Long, *The Agricultural Revolution*, Annotated edition. San Diego: Lucent Books, 2004.

[10] M. Overton, *Agricultural Revolution in England: The Transformation of the Agrarian Economy 1500-1850*. Cambridge; New York: Cambridge University Press, 1996.

[11] P. N. Stearns, *The Industrial Revolution in World History*, Fourth Edition, Fourth Edition edition. Boulder, Colo: Westview Press, 2012.

[12] E. Brynjolfsson and A. McAfee, *The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies*, 1 edition. New York: W. W. Norton & Company, 2014.

[13] C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication*, First Edition. Urbana: University of Illinois Press.

[14] E. F. Codd, "A Relational Model of Data for Large Shared Data Banks," *Commun ACM*, vol. 13, no. 6, pp. 377–387, Jun. 1970.

[15] R. Ramakrishnan and J. Gehrke, *Database Management Systems, 3rd Edition*, 3rd edition. Boston: McGraw-Hill, 2002.

[16] F. Provost and T. Fawcett, *Data Science for Business: What you need to know about data mining and data-analytic thinking*, 1 edition. Sebastopol, Calif.: O'Reilly Media, 2013.

[17] S. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*, 3 edition. Upper Saddle River: Prentice Hall, 2009.

[18] A. Newell and H. Simon, “Computer Science as Empirical Inquiry: Symbols and Search,” *Commun. ACM*, vol. 19, no. 3, 1976.

[19] J. R. Anderson, *Cognitive Psychology and Its Implications*. New York: Worth Publishing, 1999.

[20] P. Thagard, *Computational Philosophy of Science*. Cambridge, MA: MIT Press, 1988.

[21] H. P. Stapp, *Mindful Universe: Quantum Mechanics and the Participating Observer*, 2nd ed. 2011 edition. Berlin; New York: Springer, 2011.

[22] W. R. Gilks, S. Richardson, and D. Spiegelhalter, *Markov Chain Monte Carlo in Practice*. London: Chapman & Hall, 1996.

[23] D. H. Ackley, G. E. Hinton, and T. J. Sejnowski, “A Learning Algorithm for Boltzmann Machines*,” *Cogn. Sci.*, vol. 9, no. 1, pp. 147–169, Jan. 1985.

[24] P. Diaconis, “The Markov chain Monte Carlo revolution,” *Bull. Am. Math. Soc.*, vol. 46, no. 2, pp. 179–205, 2009.

[25] W. Heisenberg, “The Representation of Nature in Contemporary Physics,” *Daedalus*, vol. 87, pp. 95–108, 1958.

[26] J. von Neumann, “Mathematical Foundations of Quantum Mechanics (Chapter VI),” Princeton, NJ: Princeton University Press., 1932.

[27] H. P. Stapp, “Mind, brain and neuroscience,” *BIONOETICS*, vol. 10, no. 1, pp. 227-231, 2014.

[28] R. M. Neal, “MCMC using Hamiltonian dynamics,” *ArXiv12061901 Phys. Stat*, Jun. 2012.