

The Cognitive Cycle

John F. Sowa

Abstract. In the twenty years from first grade to a PhD, students never learn any subject by the methods for which machine-learning algorithms have been designed. Those algorithms are useful for analyzing large volumes of data. But they don't enable a computer system to learn a language as quickly and accurately as a three-year-old child. They're not even as effective as a mother raccoon teaching her babies how to find the best garbage cans. For all animals, learning is integrated with the cognitive cycle from perception to purposeful action. Many algorithms are needed to support that cycle. But an intelligent system must be more than a collection of algorithms. It must integrate them in a cognitive cycle of perception, learning, reasoning, and action. That cycle is key to designing intelligent systems.

1. Theories of Learning and Reasoning

The nature of the knowledge stored in our heads has major implications for educating children and for designing intelligent systems. Both fields organize knowledge in teachable modules that are presented in textbooks and stored in well structured databases and knowledge bases. A systematic organization makes knowledge easier to teach and to implement in computer systems. But as every student discovers upon entering the workforce, "book learning" is limited by the inevitable complexities, exceptions, and ambiguities of engineering, business, politics, and life. Although precise definitions and specifications are essential for solving problems in mathematics, science, and engineering, most problems aren't well defined. As Shakespeare observed, "There are more things in heaven and earth, Horatio, than are dreamt of in your philosophy."

During the past half century, neuroscience has discovered a great deal about the organization of the human brain and its inner workings. But each discovery has led to far more questions than answers. Meanwhile, artificial intelligence developed theories, tools, and algorithms that have been successfully applied to practical applications. Both neuroscience and AI have been guided by the centuries of research in other branches of cognitive science: philosophy, psychology, linguistics, and anthropology.

One explanation of learning explanation of learning that has been invented and reinvented in various guises is the *apperceptive mass* or dominant system of ideas. It was suggested by Leibniz, elaborated by Johann Friederich Herbart [5], and had a strong influence on educational psychology [31]:

There is a unity of consciousness — attention, as one might call it — so that one cannot attend to two ideas at once except in so far as they will unite into a single complex idea. When one idea is at the focus of the consciousness it forces incongruous ideas into the background or out of consciousness altogether. Combined ideas form wholes and a combination of related ideas form an apperceptive mass, into which relevant ideas are welcomed but irrelevant ones are excluded.... If information is to be acquired as easily and as rapidly as possible, it follows that in teaching one should introduce new material by building upon the apperceptive mass of already familiar ideas. Relevant ideas, then, will be most easily assimilated to the apperceptive mass, while irrelevant ideas will tend to be resisted and, consequently, will not be assimilated as readily.

In AI, versions of semantic networks from the 1960s [20] to the 1990s [8] resemble Herbart's apperceptive mass. In fact, the implementations exhibit many of the properties claimed by educational psychologists [26]. Piaget and his colleagues in Geneva observed children and analyzed the *schemata* they used at various ages [19]. They showed that learning progresses by stages, not by a series of database updates. At each stage, the brain *assimilates* new information to its existing structures. When the complexity of the information grows beyond the capacity of the structures at one stage, a minor revolution occurs, and a new schema is created to reorganize the mental structures. The later, more abstract *conceptual* schemata are derived by generalizing, building upon, and reorganizing the *sensorimotor* schemata formed in infancy and early childhood.

Rumelhart and Norman [23] used the term *accretion* for Piaget's assimilation of new information to the old schemata. They split Piaget's notion of *accommodation* in two stages of *tuning* and *restructuring*:

- *Accretion*. New knowledge may be added to episodic memory without changing semantic memory. It corresponds to database updates that state new facts in terms of existing categories and schemata.
- *Schema tuning*. Minor changes to semantic memory may add new features and categories, generalize old schemata to more general supertypes, and revise priorities or defaults.
- *Restructuring*. As episodic memory becomes more complex, a major reorganization of semantic memory may be needed. Completely new schemata and concept types may represent old information more efficiently, support complex deductions in fewer steps, and make drastic revisions to old ways of thinking and acting.

Restructuring is responsible for the *plateau effect*: people quickly learn a new skill up to a modest level of proficiency; then they go through a period of stagnation when study and practice fail to show a noticeable improvement; suddenly, they break through to a new plateau where they again progress at a rapid rate — until they reach the next plateau. Restructuring is the creative discovery. When a person attains a new insight, a sudden revolution reorganizes the old information in the new categories and schemata.

The paradigms that psychologists have proposed for human learning have their counterparts in AI. In 1983, a review of machine learning [3], distinguished three historical periods, each characterized by its own paradigm for learning:

- *Stimulus-response*. In the late 1950s and early 60s, *neural nets* and *self-organizing systems* were designed to start from a neutral state and build up internal connections solely by reinforcement and conditioning.
- *Induction of concepts*. A higher form of learning is the induction of new types of categories from the data. It started in the 1960s with clustering and concept learning techniques, and it can be integrated with every learning system, formal or informal.
- *Knowledge-intensive learning*. Before a system can learn anything new, it must already have a great deal of initial knowledge. Most methods of restructuring are compatible with knowledge-intensive learning.

One of the first steps in restructuring is categorization: selecting new concept types, mapping percepts to those types, and assigning the types to appropriate levels of the type hierarchy. A more radical alteration of schemata is a leap into the unknown. It requires someone to abandon comfortable old ways of thought before there is any reassurance that a new schema is better. To explain how such learning is possible, Charles Sanders Peirce [17] proposed the reasoning method of *abduction*, which operates in a *cognitive cycle* with observation, induction, abduction, deduction, testing, action — and

repeat:

- Deduction is logical inference by formal reasoning or plausible reasoning.
- Induction involves gathering and generalizing new data according to existing types and schemata.
- Abduction consists of a tentative guess that introduces a new type or schema, followed by deduction for exploring its consequences and induction for testing it with reality.

An abduction is a hunch that may be a brilliant breakthrough or a dead end. The challenge is to develop a computable method for generating insightful guesses. Beforehand, a guess cannot be justified logically or empirically. Afterwards, its implications can be derived by deduction and be tested against the facts by induction. Peirce's logic of pragmatism integrates the reasoning methods with the twin gates of perception and purposeful action.

2. Deep Learning

Educational psychologists, who were strongly influenced by Herbart, Piaget, and William James, distinguish *deep learning* as a search for meaning from *surface learning* as the memorization of facts. Although their definitions are not sufficiently precise to be implemented in AI systems, surface learning corresponds to accretion, and deep learning corresponds to schema tuning and restructuring.

Educational psychologists considered behaviorism *rat psychology*. But Thorndike [29], a former student of William James, used animal experiments to develop a stimulus-response theory, which he called *connectionism*: rewards strengthen the S-R connections, and punishments weaken them. In 1943, McCulloch and Pitts [13], a neurophysiologist collaborating with a logician, designed a theoretical model of neural nets that were capable of learning and computing any Boolean function of their inputs. To implement a version, Rosenblatt used vacuum tubes to simulate the nodes of a neural net he called a *perceptron* [22]. Later neural nets were implemented by programs on a digital computer.

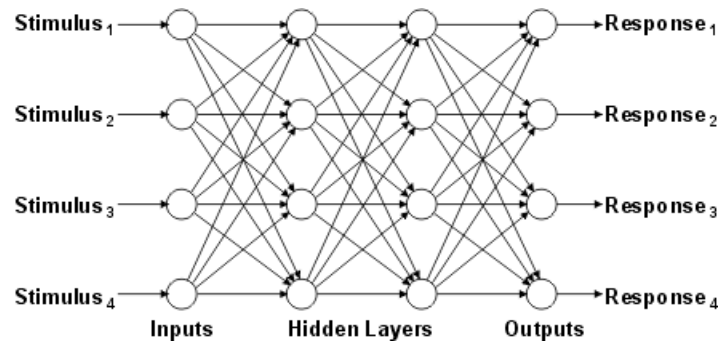


Figure 1: A neural net for connecting stimuli to responses

Behaviorist methods of operant conditioning suggested neural-net methods of learning by *backpropagation*. Figure 1 shows a neural net with stimuli entering the layer of nodes at the left. The nodes represent neurons, and the links represent axons and dendrites that propagate signals from left to right. In computer simulations, each node computes its output values as a function of its inputs. Whenever a network generates an incorrect response to the training stimuli, it is, in effect, “punished” by adjusting weights on the links, starting from the response layer and propagating the adjustments backwards toward the stimulus layer.

Figure 1 is an example of a multilayer feedforward neural network. The original perceptrons had just a single layer of nodes that connected inputs to outputs. With more layers and the option of cycles that provide feedback from later to earlier nodes, more complex patterns can be learned and recognized. After a network has been trained, it can rapidly recognize and classify patterns. But the training time for backpropagation increases rapidly with the number of inputs and the number of hidden layers.

Many variations of algorithms and network topologies have been explored, but the training time for a network grew exponentially with the number of layers. A major breakthrough for *deep belief networks* was a strategy that reduced the training time by orders of magnitude. According to Hinton [6], the key innovation was “an efficient, layer by layer procedure” for determining how the variables at one layer depend on the variables at the previous layer.

Deep belief nets are learned one layer at a time by treating the values of the latent variables in one layer, when they are being inferred from data, as the data for training the next layer. This efficient, greedy learning can be followed by, or combined with, other learning procedures that fine-tune all of the weights to improve the generative or discriminative performance of the whole network.

With this strategy, the time for training a network with N layers depends on the sum of the times for each layer, not their product. As a result, deep belief nets can quickly learn and recognize individual objects, even human faces, in complex street scenes.

The adjective *deep* in front of *belief network* introduced an ambiguity. The term *deep learning*, which the educational psychologists had used for years, implies a human level of understanding that goes far beyond pattern recognition. Enthusiastic partisans of the new methods for training neural networks created an ambiguity by adopting that term. To avoid confusion, others use the more precise term *deep neural network* (DNN). The remainder of this article will use the acronym DNN for methods that use only the neural networks. For hybrid systems that combine a DNN with other technologies, the additional methods and the roles they play are cited.

The Stanford NLP group led by Christopher Manning has been in the forefront of applying statistical methods to language analysis. By combining DNNs for scene recognition with statistical NLP methods, they relate objects in a scene to sentences that describe the scene — for example, the sentence *A small crowd quietly enters the historic church* and the corresponding parts of the scene [25]. But to derive implications of the sentences or the scenes, they use a variation of logic-based methods that have been used in AI for over 40 years [9]. Some of those methods are as old as Aristotle and Euclid. Others were developed by Peirce and Polya [28].

DNNs are highly successful for static pattern recognition. Other techniques, such as hidden Markov models (HMMs), are widely used for recognizing time-varying sequences. For a dissertation under Hinton’s supervision, Navdeep Jaitly developed a DNN-HMM hybrid for speech recognition [7]. His office mate, Volodymyr Mnih, combined DNNs with a technique, called *Q-learning*, which uses patterns from two or more time steps of a neural net to make HMM-like predictions [30]. By using DNNs with Q-learning, Mnih and his colleagues at DeepMind Technologies designed the DQN system, which learned to play seven video games for the Atari 2600: Pong, Breakout, Space Invaders, Seaquest, Beamrider, Enduro, and Q*bert [15]. DQN had no prior information about the objects, actions, features, or rules of the games. For each game, the input layer receives a sequence of screen shots: 210×160 pixels and the game score at each step. Each layer learns features, which represent the input data for the next layer. The final layer determines which move to make for the next time step based on patterns in the current and previous time steps.

When compared with other computer systems, DQN outperformed all other machine learning methods on 6 of the 7 games. Its performance was better than a human expert on Breakout, Enduro, and Pong.

Its score was close to human performance on Beamrider. But Q*bert, Seaquest, and Space Invaders require long-term strategy. On those games, DQN performed much worse than human experts. Yet the results were good enough for Google to buy the DeepMind company for \$400 million [21].

In a critique of DNNs [11], the psycholinguist Gary Marcus wrote “There is good reason to be excited about deep learning, a sophisticated machine learning algorithm that far exceeds many of its predecessors in its abilities to recognize syllables and images. But there’s also good reason to be skeptical... deep learning takes us, at best, only a small step toward the creation of truly intelligent machines.”

Advocates of DNNs claim that they embody “a unified theory of the human cortex” that holds the key to all aspects of intelligence. But Marcus observed “They have no obvious ways of performing logical inferences, and they are also still a long way from integrating abstract knowledge, such as information about what objects are, what they are for, and how they are typically used. The most powerful AI systems, like Watson, the machine that beat humans in Jeopardy!, use techniques like deep learning as just one element in a very complicated ensemble of techniques.” After a discussion with Peter Norvig, director of Google Research, Marcus reported that “Norvig didn’t see how you could build a machine that could understand stories using deep learning alone.”

3. Models and Reality

Language, thought, and logic are systems of signs. They are related to the world (or reality), but in different ways. Humans and other animals relate their internal signs (thoughts and feelings) to the world by perception and purposeful action. They also use their sensorimotor organs for communicating with other animals by a wide range of signs, of which the most elaborate are human languages, natural or artificial. Cognitive scientists — philosophers, psychologists, linguists, anthropologists, neuroscientists, and AI researchers — have used these languages to construct an open-ended variety of models and theories about these issues.

As Charles Sanders Peirce observed, all the models, formal and informal, have only two things in common: first, they consist of signs about signs about signs...; second, they are, at best, fallible approximations to reality. As engineers say, all models are wrong, but some are useful. To illustrate the relationships, Figure 2 shows a model as a Janus-like structure, with an engineering side facing the world and an abstract side facing a theory.

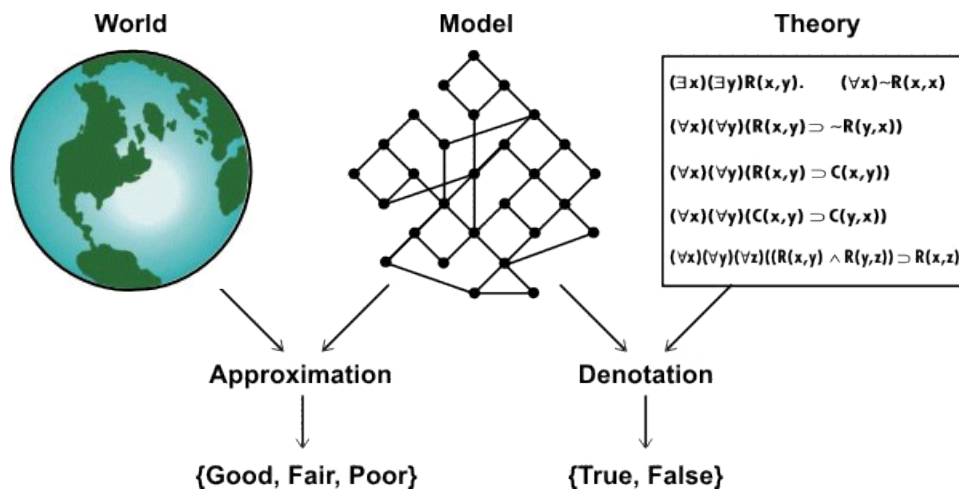


Figure 2: A model that relates a theory to the world

On the left is a picture of the physical world, which contains more detail and complexity than any humanly conceivable model or theory could represent. In the middle is a mathematical model that represents a domain of individuals \mathbf{D} and a set of relations \mathbf{R} over individuals in \mathbf{D} . If the world had a unique decomposition into discrete objects and relations, the world itself would be a universal model, of which all correct models would be subsets. But the selection of a domain and its decomposition into objects depend on the intentions of some agent and the limitations of the agent's measuring instruments. Even the best models are approximations to a limited aspect of the world for a specific purpose.

The two-stage mapping from theories to models to the world can reconcile a Tarski-style model theory with the fuzzy methods of Lotfi Zadeh [32]. In Tarski's models, each sentence has only two possible truth values: $\{T,F\}$. In fuzzy logic, a sentence can have a continuous range of values from 0.0 for certainly false to 1.0 for certainly true. Hedging terms, such as likely, unlikely, very nearly true, or almost certainly false, represent intermediate values. The two-stage mapping of Figure 2 can accommodate an open-ended variety of models and methods of reasoning. In addition to the Tarski-style models and the fuzzy approximations, it can represent engineering models that use any mathematical, computational, or physical methods for simulating a physical system. For robot guidance, the model may represent the robot's current environment or its future goals. For mental models, it could represent the virtual reality that people have in their heads.

No discrete model can be an exact representation of a physical system. But discrete models can exactly represent a digital computation or a mathematical game such as chess, checkers, or go. In fact, the game of checkers has a far deeper strategy than any of those Atari games. But in 1959, Art Samuel wrote a program that learned to play checkers better than he could [24]. He ran it on an IBM 704, a vacuum-tube computer that had only 144K bytes of RAM and a CPU that was much slower than the Atari. Yet it won one game in a match with the Connecticut state checkers champion.

For the learning method, Samuel used a weighted sum of features to evaluate each position in a game. His algorithm was equivalent to a one-layer neural net. But the program also tested each evaluation by looking ahead several moves. During a game, it maintained an exact model of each state of the game. By searching a few moves ahead, it would determine exact sequences, not the probabilities predicted by an HMM. In the match with the human expert, Samuel's program found the winning move when the human made a mistake.

Samuel's program was a hybrid of a statistical learning method with an exact model for positions in the game. Many chess-playing programs use a similar hybrid, but the programs that model game positions are completely different. Some AI researchers have claimed that to exhibit a human level of intelligence, a *general game playing system* should be able to learn multiple games without special programming for each one [4]. But the criterion of "special programming" is unclear:

- The rules for games of strategy, such as bridge, chess, and go can be learned in a day. But mastery requires years.
- Many people have become good amateur players of all three games, but no one has ever reached a world-class mastery of more than one. Grandmasters in those games begin early in life, preferably before puberty, and devote many hours per week to reach a world-class level.
- Early learning is also necessary for native mastery of languages, musical performance, gymnastics, and other complex skills. Mentoring by a master is also critical. That concentrated study could be considered a kind of special programming.

4. The Cognitive Cycle

The human brain is a complex hybrid of multiple components with different kinds of representations interacting at various stages during the processes of learning and reasoning. After many years of research in the design and implementation of AI systems, Minsky [14] argued that no single mechanism, by itself, can adequately support the full range of functions required for a human level of intelligence:

What magical trick makes us intelligent? The trick is that there is no trick. The power of intelligence stems from our vast diversity, not from any single, perfect principle. Our species has evolved many effective although imperfect methods, and each of us individually develops more on our own. Eventually, very few of our actions and decisions come to depend on any single mechanism. Instead, they emerge from conflicts and negotiations among societies of processes that constantly challenge one another.

Evidence from fMRI scans supports Minsky's hypothesis. Mason and Just [12] studied 14 participants who were learning the internal mechanisms of four devices: a bathroom scale, a fire extinguisher, an automobile braking system, and a trumpet. For all 14 participants, the brain regions involved in learning how each device worked progressed through the same stages:

(1) initially, the representation was primarily visual (occipital cortex); (2) it subsequently included a large parietal component; (3) it eventually became cortically diverse (frontal, parietal, temporal, and medial frontal regions); and (4) at the end, it demonstrated a strong frontal cortex weighting (frontal and motor regions). At each stage of knowledge, it was possible for a classifier to identify which one of four mechanical systems a participant was thinking about, based on their brain activation patterns.

In the first stage, the visual cortex recognized and encoded the visual forms and details of each device. In the second stage, the parietal lobes, which represent cognitive maps or schemata, were involved in "imagining the components moving." The third stage involved all lobes of the cortex. Since the medial frontal cortex has motor connections to all parts of the body, the participants were probably "generating causal hypotheses." Finally, the frontal cortex was anticipating "how a person (probably oneself) would interact with the system."

An educational psychologist would not find the studies by Mason and Just to be surprising. But a partisan of DNNs might find them disappointing. DNNs may be useful for simulating some aspects of intelligence, but AI and cognitive science have developed many other useful components. In the book *Deep Learning*, the psychologist Stellan Ohlsson reviewed those developments and systematic ways of integrating them [16]. Ohlsson defined learning as "nonmonotonic cognitive change," which "overrides experience" by

- creating novel structures that are incompatible with previous versions,
- adapting cognitive skills to changing circumstances, and
- testing those skills by acting upon the environment.

Ohlsson cites Peirce's logic of pragmatism [17] and adopts Peirce's version of abduction as the key to nonmonotonic reasoning. The result is similar to Peirce's cycle of pragmatism (Figure 3).

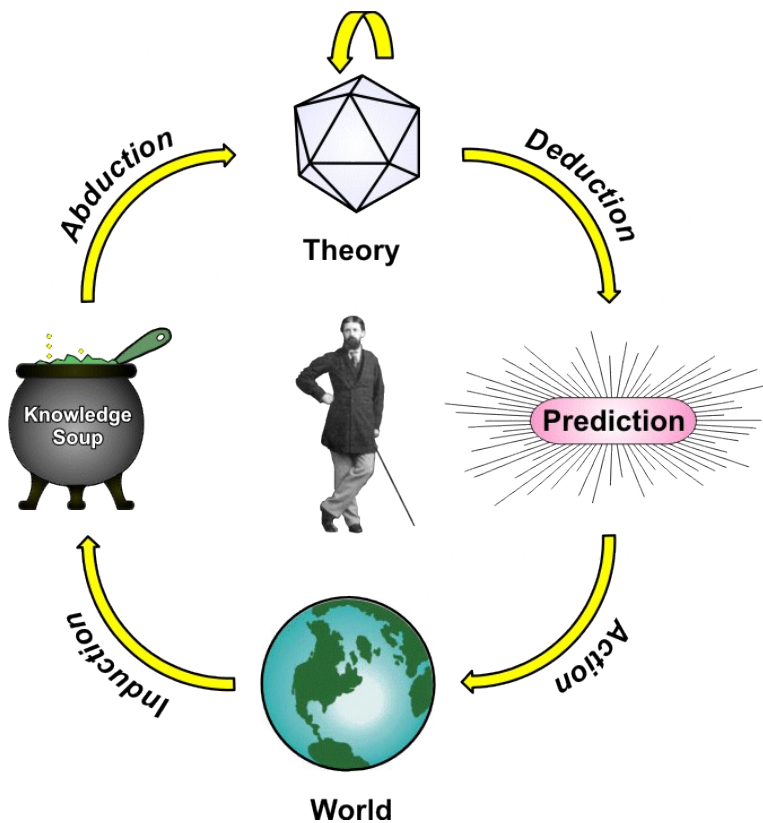


Figure 3: Peirce's cycle of pragmatism

In Figure 3, the pot of knowledge soup represents the highly fluid, loosely organized accumulation of memories, thoughts, fantasies, hopes, and fears in the mind. The arrow of induction represents new observations and generalizations that are tossed in the pot. The crystal at the top symbolizes the elegant, but fragile theories that are constructed by abduction from chunks in the knowledge soup. The arrow above the crystal indicates the process of belief revision, which uses repeated abductions to modify the theories. At the right is a prediction derived from a theory by deduction. That prediction leads to actions whose observable effects may confirm or refute the theory. Those observations are the basis for new inductions, and the cycle continues.

Learning is the process of accumulating chunks of knowledge in the soup and organizing them into theories — collections of consistent beliefs that prove their value by making predictions that lead to successful actions. Learning by any agent — human, animal, or robot — involves a constant cycling from data to models to theories and back to a reinterpretation of the old data in terms of new models and theories. Beneath it all, there is a real world, which the entire community of inquirers learns to approximate through repeated cycles of induction, abduction, deduction, and action.

The cognitive cycle is not a specific technology, such as a DNN, an HMM, or an inference engine. It's a framework for designing hybrid systems that can accommodate multiple components of any kind. Ohlsson, for example, is a psychologist who had collaborated with AI researchers, and he was inspired by Peirce's writings to design systems based on a cognitive cycle that is similar to Figure 3. James Albus was an engineer who studied neuroscience to get ideas for designing robots [1]. Although he had not studied Peirce, he converged on a similar cycle. Majumdar and Sowa [10] did study Peirce, and they adopted Figure 3 as their foundation. Each of the five arrows in Figure 3 may be implemented in a variety of different methods, formal or informal, crisp or fuzzy, statistical or symbolic, declarative or procedural.

In the conclusion of an article about statistical modeling, the statistician Leo Breiman stated an important warning [2]:

Oddly, we are in a period where there has never been such a wealth of new statistical problems and sources of data. The danger is that if we define the boundaries of our field in terms of familiar tools and familiar problems, we will fail to grasp the new opportunities.

Another statistician, Martin Wilk [33], observed “The hallmark of good science is that it uses models and ‘theory’ but never believes them.” The logician Alfred North Whitehead stated a similar warning about grand theories of any kind [32]: “Systems, scientific and philosophic, come and go. Each method of limited understanding is at length exhausted. In its prime each system is a triumphant success: in its decay it is an obstructive nuisance.”

The cognitive cycle is a metalevel framework. It relates the methods of reasoning that people use in everyday life and scientific research. The cycle is self correcting. Every prediction derived in one cycle is tested in the next cycle. Although perfect knowledge is an unattainable goal, repeated cycles can converge to knowledge that is adequate for the purpose. That is the logic of pragmatism.

References

1. Albus, James S. (2010) A model of computation and representation in the brain, *Information Sciences* **180**, 1519–1554. <http://www.james-albus.org/docs/ModelofComputation.pdf>
2. Breiman, Leo (2001) [Statistical Modeling: The Two Cultures](#), *Statistical Science* **16:3**, 199-231.
3. Carbonell, Jaime G., Ryszard S. Michalski, & Tom M. Mitchell (1983) An overview of machine learning, in Michalski, Carbonell, & Mitchell, *Machine Learning*, Palo Alto: Tioga, pp. 3-23.
4. Genesereth, Michael, Nathaniel Love, & Barney Pell (2005) General Game Playing: Overview of the AAAI Competition, *AI Magazine* **26:2**, 62-72.
5. Herbart, Johann Friedrich (1816) *A Textbook in Psychology*, translated by M. K. Smith, New York: Appleton, 1891.
6. Hinton, Geoffrey E. (2009) Deep belief networks, *Scholarpedia*, 4(5):5947, http://www.scholarpedia.org/article/Deep_belief_networks
7. Jaitly, Navdeep (2014) Exploring deep learning methods for discovering features in speech recognition, PhD Dissertation, University of Toronto. http://www.cs.toronto.edu/~ndjaitly/Jaitly_Navdeep_201411_PhD_thesis.pdf
8. Lamb, Sydney M. (1999) *Pathways of the Brain: The Neurocognitive Basis of Language*, Amsterdam: John Benjamins.
9. MacCartney, Bill, & Christopher D. Manning (2014) Natural logic and natural language inference, in Harry Bunt et al., *Computing Meaning* **4**, Berlin: Springer, pp. 129–147.
10. Majumdar, Arun K., & John F. Sowa (2009) Two paradigms are better than one and multiple paradigms are even better, in S. Rudolph, F. Dau, and S.O. Kuznetsov, eds., *Proceedings of ICCS'09*, LNAI 5662, Springer, pp. 32-47. <http://www.jfsowa.com/pubs/paradigm.pdf>
11. Marcus, Gary F. (2012) Is “deep learning” a revolution in artificial intelligence? *New Yorker*, 25 November 2012. <http://www.newyorker.com/news/news-desk/is-deep-learning-a-revolution-in-artificial-intelligence>
12. Mason, Robert A., & Marcel Adam Just (2015) Physics instruction induces changes in neural knowledge representation during successive stages of learning, *NeuroImage* **111**, 36-48.
13. McCulloch, Warren S., & Walter Pitts (1943) A logical calculus of the ideas immanent in nervous activity, *Bulletin of Mathematical Biophysics* **5**, 115-133.

14. Minsky, Marvin (1986) *The Society of Mind*, New York: Simon & Schuster, Section 30.8.
15. Mnih, V., K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra, M. Riedmille (2013) Playing Atari with deep reinforcement learning, NIPS Deep Learning Workshop, <https://www.cs.toronto.edu/~vmnih/docs/dqn.pdf>
16. Ohlsson, Stellan (2011) *Deep Learning: How the Mind Overrides Experience*, Cambridge: University Press.
17. Peirce, Charles Sanders (1903) *Pragmatism as a Principle and Method of Right Thinking*, The 1903 Lectures on Pragmatism, ed. by P. A. Turrissi, SUNY Press, Albany, 1997. Also in [18], pp. 131-241.
18. Peirce, Charles Sanders (EP) *The Essential Peirce*, ed. by N. Houser, C. Kloesel, and members of the Peirce Edition Project, 2 vols., Indiana University Press, Bloomington, 1991-1998. Peirce, Charles Sanders (CP) *Collected Papers of C. S. Peirce*, ed. by C. Hartshorne, P. Weiss, & A. Burks, 8 vols., Cambridge, MA: Harvard University Press, 1931-1958.
19. Piaget, Jean (1968) *On the Development of Memory and Identity*, Barre, MA: Clark University Press.
20. Quillian, M. Ross (1968) Semantic memory, in M. Minsky, *Semantic Information Processing*, Cambridge, MA: MIT Press, pp. 227-270.
21. Regalado, Antonio (2014) Is Google cornering the market on deep learning? *Technology Review*, 29 January 2014. <http://www.technologyreview.com/news/524026/is-google-cornering-the-market-on-deep-learning/>
22. Rosenblatt, Frank (1958) The perceptron: a probabilistic model for information storage and organization in the brain, *Psychological Review* **65:6**, 386-408.
23. Rumelhart, David E., & Donald A. Norman (1978) Accretion, tuning, and restructuring: three modes of learning, in J. W. Cotton & R. L. Klatzky, eds., *Semantic Factors in Cognition*, Hillsdale, NJ: Lawrence Erlbaum, pp. 37-54.
24. Samuel, Arthur L. (1959) Some studies in machine learning using the game of checkers, *IBM Journal of Research and Development* **3**, 211-229.
25. Socher, Richard, Cliff Chiung-Yu Lin, Andrew Y. Ng, & Christopher D. Manning (2011) Parsing natural scenes and natural language with recursive neural networks, ICML 2011. http://www.socher.org/uploads/Main/SocherLinNgManning_ICML2011.pdf
26. Sowa, John F. (1992) Semantic networks, *Encyclopedia of Artificial Intelligence*, Second Edition, edited by S. C. Shapiro, Wiley, New York. Updated version at <http://www.jfsowa.com/pubs/semnet.pdf>
27. Sowa, John F. (2006) The challenge of knowledge soup, in J. Ramadas & S. Chunawala, *Research Trends in Science, Technology, and Mathematics Education*, Mumbai: Homi Bhabha Centre, pp. 55-90. <http://www.jfsowa.com/pubs/challenge.pdf>
28. Sowa, John F. (2015) Peirce, Polya, and Euclid: integrating logic, heuristics, and geometry, lecture presented at the American Philosophical Association conference, Vancouver, April 2015. Slides at <http://www.jfsowa.com/talks/ppe.pdf>
29. Thorndike, Edward Lee (1932) *The Fundamentals of Learning*, New York: Teachers College Press.
30. Watkins, Christopher J.C.H., & Peter Dayan (1992) Q-learning, *Machine Learning* **8:3-4**, 279-292.
31. Watson, Robert I. (1963) *The Great Psychologists from Aristotle to Freud*, New York: Lippencot, pp. 209-210.
32. Whitehead, Alfred North (1933) *Adventures of Ideas*, New York: Macmillan.
33. Wilk, Martin, as quoted by John W. Tukey (1962) The future of data analysis, *Annals of Mathematical Statistics* **33:1**, 1-67.
34. Zadeh, Lotfi A. (1965) Fuzzy sets, *Information and Control* **8**, 338-353.