

## **The Cup Whisperer.** Review of

A Thousand Brains: A New Theory of Intelligence Hardcover – March 2, 2021

by Jeff Hawkins (Author), Richard Dawkins (Foreword)

Review for the American Journal of Psychology. Review by Bill Rowe.

wrowe@ucsc.edu

Once upon a time about five hundred million years ago a tadpole look-alike creature swam the ocean depths. It was the juvenal phase of an ancient sea squirt, sometimes referred to as a tunicate. Its job at this developmental stage was to find a suitable rock on the ocean floor to anchor down upon and morph into an adult sea squirt. This adult form would spend the rest of its life attached to that rock, passively feeding on what passing currents would bring its way. Not a bad life. But that's not the end of our story, it's just the beginning. Because a very strange twist occurs in this tale. Some say it's apocryphal, others say an exaggeration, but I think it might be true. Give or take.

One day an enterprising juvenile tunicate decided not to settle down on a rock, it just kept swimming. And swimming. It got bigger, and bigger. It developed eyes and gills. And got bigger still. It developed fins and sexual reproduction. It got bigger and fins developed into feet, and it walked upon the land trading gills for lungs. It spent time walking on all fours and time living in trees. It came down from the trees and walked upright and learned to talk to other decedents of juvenile tunicates. And then one day it looked back out upon the ocean and said, "thanks." Thanks to that enterprising juvenile tunicate who decided to not eat its own nervous system. Wait, what! Not do what?

Wasn't she supposed to say something like "thanks oh ancient one for your bold and courageous swimming" something like that? Not really. If it were that simple there would be no introduction like this for a review of a book like A Thousand Brains. As I mentioned earlier, there's a twist to this tale.

The twist comes in the strange turn of events following the typical juvenile tunicate's successful rock anchoring. In the words of the British Neuroscientist Daniel Wolpert, "It digests its own nervous system for food!" He says this in the opening to his 2011 TED talk on the topic of why do we have brains. And he explains that we have

brains for one reason and one reason only, and that is to have adaptable and complex movements, there is no other reason to have a brain. You see, even tadpole-style swimming requires a brain. Not a lot of brain. A rudimentary nervous system like that of the juvenile tunicate gets the job done. But once you don't need to move, Wolpert says, you don't need that brain. Thus the tunicate's first meal.

Hawkins doesn't reference Wolpert in his book (Although Wolpert is mentioned in several Numenta papers.) But they are kindred spirits. Wolpert by his own declaration is a movement chauvinist. Hawkins, by his inspired dedication, is an evangelist. As you will see when you read the book "to be" is to be moving. That's my way of capturing the defining feature of Hawkins' model. We have to physically "graze" the world in order to perceptually grasp it. The story Hawkins weaves is compelling. But all good stories yearn for an origin, a genesis. So I recruited J.T., the Juvenal Tunicate. Perhaps a mythical, but in any case a likeable character. An ancestor who refused to sit on a rock and eat its own brain. An ancestor who used that rudimentary nervous system to set out across deep waters, swimming unknowns and surfing uncertainties. Thanks J.T.

## A Thousand Brains

The full title of Hawkins's book is A Thousand Brains. A New Theory of Intelligence. There are 263 pages sectioned into three parts - A New Understanding of the Brain, Machine Intelligence, and Human Intelligence - for a total of 16 chapters and a summary of Final Thoughts, all introduced with a Forward by none other than Richard Dawkins. Dawkins does not hold-back his praise. He opens his commentary by noting parallels between Charles Darwin and Jeff Hawkins. Yes, *that* Charles Darwin. One parallel being the fact that Darwin did work, and Hawkins is currently doing work, outside of universities and without government research grants. "Well, you get the parallel," says Dawkins. Yes. And clearly knowing that that's a lot of parallels to live up to Dawkins goes on to say that the ideas of both men require book length treatments. He gets specific about this by calling attention to *reference frames* which are key features in Hawkins's theory. These reference frames, as we learn in the book, are used to make predictions and orchestrate movements. And taking reference frames even further, "The very act of *thinking* is a form of movement." (Hawkins, 2021, p. vii). "Bulls-

eye!" Dawkins exclaims, the ideas of both men are enough to fill a book. He ends the introduction by noting how a book revealing that the brain works in such a way is "nothing short of exhilarating."

So, will the Bay Area really be the next Galapagos? And brighter-than-a-bird's-brain computer chips be Hawkins's Finches? History will decide. But in the meantime it is just possible that Hawkins is enough outside the constraints of normal science to pull together threads that are not normally found in the same fabric. Keep in mind that this thing we call thinking has been elusive for a long time. Long enough and elusive enough to drive many-a-thinker into "get thee to a nunnery" exasperation. In this case the "nunnery" is a place just outside of Plato's cave. The place where pure things exist. "Pure" being a stand-in for stuff we can't explain.

But maybe we can explain thinking. Maybe the features of thinking are just the same features as planning, predicting, executing, and verifying movements. And just maybe a full court press at trying to build a machine to emulate that citadel of thought, the neocortex, will shine new light on an old problem. I'm reviewing this book because I think that is worth a shot.

## Part 1, A New Understanding of the Brain.

This first section is part personal story and part technical description. The personal part, initially to understand a framework for how the brain works and, subsequently, to use that understanding to re-conceptualize computers, began in 1979. That's when Hawkins read a Scientific American article by the molecular biologist Francis Crick called "Thinking About the Brain." In that article Crick called attention to the fact that there was a large quantity of facts that scientists had collected about the brain. But in spite of this knowledge accumulation the brain's workings were still mysterious and, further, there was a conspicuous absence of any sort of broad unifying framework of ideas. The young Hawkins was inspired by Crick's essay. He felt that this mystery could be solved in his lifetime and that is exactly what he set out to do.

The technical part is woven in and around this personal story. Fundamentally it recasts the brain as an organ of prediction. As Hawkins and Blakeslee say in a 2016 paper, "We propose that the most fundamental operation of all neocortical tissue is

learning and recalling sequences of patterns." And they evoke the sentiments of Karl Lashley that this is "the most important and also the most neglected problem of cerebral physiology" (Lashley, 1951). Hawkins takes us through this understanding of the brain in chapters 1 through 7. We learn that the brain is doing one thing over and over again; it is forming models of the world by sampling it sequentially. This is basically what Hawkins's "New Theory of Intelligence" is about. Intelligence is the ability to form models of the world. One thing that is important to understand here is that this definition of intelligence is independent of goals or drives. Hawkins is not trying to reverse engineer a person. So intelligence is like a map. It is a tool for achieving a goal, but it has no desires or aspirations of its own. Later in this review we will look at the details of how sequence memory is acquired at learning time and at how prediction is initiated at remembering time.

Hawkins attended Cornell university where he received a bachelors degree in electrical engineering in 1979. After that he worked for a short time at Intel but moved on in 1982 to a smaller more agile company called GRiD Systems. He also took some time to apply as a graduate student at MIT's A.I. lab. He was told that his proposal to create intelligent machines based on brain theory was pointless because the brain was just a messy computer. Whatever sense or no-sense that made it did not deter Hawkins. Back at GRiD he participated in pioneering many of the technologies that are present in the mobile and handheld devices that we have today.

But the work at GRiD did not offer the opportunities to explore the unifying features of the brain that were still first and foremost in his mind. In 1986 Hawkins enrolled in a neuroscience PhD program at the University of California, Berkeley. As he tells the story, his ideas and ambitions were well received. His approach to understanding the brain was viewed as sound and there was wide agreement that it was one of the most important goals in modern science. There was one problem, however, that he says he did not foresee. He was told that to get through the program he would have to work for a professor doing what that professor was doing. And there was no one at Berkeley doing what he wanted to do. Disappointment here and there never seemed to bother Hawkins all that much. So, for the next two years he took advantage of the setting by spending his days in the university's libraries getting what

he calls a "first-class albeit unconventional education." Hawkins then returned to GRiD and created one of the first tablet computers called the GridPad. The rest, as the saying goes, "is history." In 1992 Hawkins founded Palm Computing beginning a ten-year span of innovation among which are the familiar PalmPilot and the Treo. With these successes behind him, he was faced with a dilemma. Continue rolling-out ever more game-changing innovations or take a break to solve one of the world's deepest and more profound problems. He opted for the latter. Why not?

In 2002, with the help and encouragement of a few neuroscience friends, he founded the Redwood Neuroscience Institute (RNI). With 10 full-time scientists, all interested in large-scale theories of the brain, RNI became a gathering place for regular lectures open to the public, hours of discussion and debate, and during the next three years attracted over a hundred visiting scholars. But the structure of RNI did not facilitate focusing everyone's efforts on the very specific questions that motivated Hawkins. He felt he needed an organization in which he could lead his own research team. It was decided to move the Redwood Neuroscience Institute to Berkeley - yes Berkeley - where it continues today under the name The Redwood Center for Theoretical Neuroscience. Then, in 2005 Hawkins founded Numenta, an independent research company. There were two goals, to develop a theory of how the neocortex works and apply this theory about the brain to machine learning and machine intelligence. It might be a little too early to say "and the rest is history," but it is the right time to turn to the substance of Hawkins' book.

Alexa, next section, please.

### Three discoveries

It is said that a well-known British physicist once told Wolfgang Köhler that all their great discoveries came from the three B's, the Bath, the Bus, and the Bed,. This refers, no doubt, to aha moments like Archimedes' bath tub, Poincare's understanding of Fuchsian functions as his foot hit first steps of a trolley car, and August Kekulé benzene ring dream. Add to these Jeff Hawkins's coffee cup. It was a late February day in 2016, he was sitting in his office holding a Numenta coffee cup in his hands when ... well, perhaps I should let him tell the story.

"I was holding a Numenta coffee cup in my hand and observed my fingers touching it. I asked myself a simple question: What does my brain need to know to predict what my fingers will feel as they move? If one of my fingers is on the side of the cup and I move it toward the top, my brain predicts that I will feel the rounded curve of the lip. My brain makes this prediction before my finger touches the lip. What does the brain need to know to make this prediction? The answer was easy to state. The brain needs to know two things: what object it is touching (in this case the coffee cup) and where my finger will be on the cup after my finger moves. Notice that the brain needs to know where my finger is relative to the cup. It doesn't matter where my finger is relative to my body, and it doesn't matter where the cup is or how it is positioned. The cup can be tilted left or tilted right. It could be in front of me or off to the side. What matters is the location of my finger relative to the cup.

This observation means there must be neurons in the neocortex that represent the location of my finger in a reference frame that is attached to the cup. The movement-related signal we had been searching for, the signal we needed to predict the next input, was 'location on the object.' " (Hawkins, 2021. pp. 48-49).

The operant term in that quote is "reference frame." This is the Secular Grail Hawkins had been searching for since reading Crick's call for a framework back in 1979. It wasn't a stand-alone aha moment however, it was preceded by two other supporting revelations. One was in 1986 when he noticed that he could be surprised at even small changes in the locations of common objects on his desk, one of which was a cup, coincidentally. Or, if his stapler made a different sound, for example, he would notice. Likewise for the clock on the wall and the cursor on the screen. Thus he confirmed for himself what others are coming around to believing these days, and that is that the brain is an organ of prediction, and is vigilant for its violations. Or, more specifically for Hawkins, prediction is a ubiquitous function of the neocortex.

The other discovery came on the heels of answering the next logical question, "just how does a brain make these predictions." And the answer to that question required finding the commonality between predictions the brain makes under two different situations. The first one, referred to as the melody-prediction problem, concerns predictions that we make when the world changes around us. You might be

listening to a sequence of notes in a song, for example. And the other situation is about predictions we make when it is our own actions that change the world. This could be the simple sequence of reaching for a cup. These are two very different situations, yet they share an important neurological commonality; the brain is ahead of the game, forming expectations about outcomes of each sequential segment before they occur.

For the resolution to this question Hawkins drew upon the work of the Johns Hopkins neuroscientist Vernon Mountcastle. In a 1978 essay called "An Organizing Principle for Cerebral Function: The Unit Module and the Distributed System," Mountcastle had proposed that the entire neocortex was an evolutionary appendage, a change in scale from a smaller more primitive organ, but very little change in function. Meaning that the differences from region to region in the neocortex are minor compared to the similarities.

While visual experiences seem very different from tactile ones, and both of those are quite different from hearing, the physical structure of the cortical regions are much the same. The roughly 2.5mm thick neocortex is made up of six layers and those six layers are transected by functional columns that receive stimuli from small regions of the sensory organs. Those columns are, in turn, composed of smaller "mini columns" that respond to specific *features* of the stimuli coming from those small sensory regions. Mountcastle didn't specify what the function of these columns were, but his intuition was that they were all carrying out the same basic algorithm. Hawkins, inspired by Mountcastle's intuition took advantage of a Thanksgiving holiday in 2010 to figure it all out. I'll let the reader consult page 43 of a Thousand Brains for the morsels of that discovery and, for now, just relay the "take home" message that every cell in every column participates in a never-ending cycle of learning and predicting.

So thanks to a 2010 Thanksgiving holiday, Hawkins and his team had the answer to the melody-prediction problem. That understanding came 24 years after his earlier realization that the brain was basically an organ of prediction (The arc of knowledge bends slowly). But it was followed in a mere six years by a resolution to the mystery about the brain that he had hoped to resolve in his lifetime; the problem posed in Crick's 1987 paper of their being no framework uniting the many facts known about the brain.

The details of these discoveries are presented in chapter 4 called The Brain Reveals its Secrets. In his summary of that chapter Hawkins says that the goal was to introduce the reader to the idea that every cortical column in the neocortex creates reference frames. As he tells it, the resolution came to him so suddenly in his office and "I was so excited that I jumped out of my chair and ran to tell my colleague Subutai Ahmad." (Hawkins, 2021, p. 52.) In racing the 20 feet to Subutai's desk he recounts that he almost knocked over his spouse Janet who was coming to join him for lunch. But he didn't knock anyone over. Instead, his brain predicted what would happen if he carried out certain non-functional sequences and quickly organized a proper apology and generated an invitation to share a frozen yogurt. So the very process that Hawkins was studying saved a marriage, shared a yogurt, and allowed some time for seasoning before being presenting it to a friend and colleague. Sometimes things just do work out.

The three discoveries just described are detailed in a sequence of Numenta publications between 2016 and 2019. How the brain forms predictions in the presence of changing stimuli is explored in a 2016 paper called Why Neurons Have Thousands of Synapses, a Theory of Sequence Memory in Neocortex. That was followed in 2017 by a further exploration of predictive learning in the context of an active agent. That paper is called A Theory of How Columns in the Neocortex Enable Learning the Structure of the World. And the long sought after Framework in which these two predictive behaviors function is proposed in a 2019 paper called A Framework for Intelligence and Cortical Function Based on Grid Cells in the Neocortex. For details, I highly recommend reading these papers. The book provides an excellent overview of Numenta's agenda, but the papers have the facts - current models, with graphic illustrations of the neocortex, for example. Also how these models are implemented in computer designs, and discussions about the degrees successes and failures.

At at the time of this writing the mechanisms of the 2017 and 2019 papers are works in progress. They explore the complex issues of "Location," -- how the brain keeps track of the changing relationship between the sensing organ and object properties being sensed, and "Framework," -- how the brain maintains a sense of unity amidst these ever changing relationships. I mention this ongoing process of exploration and discovery only in passing since the present goal is to review Hawkins's book as it is.



That being said I encourage readers to view the videos, podcasts, and meet-ups that are made available at Numenta's website. Numenta does not hide the sometimes zig-zaggy starboard-and-port tack-and-advance process of discovery. Contrary to what von Bismark might have said, sometimes it is worthwhile to see your sausage being made. Enjoy it or not, in Numenta's case it is a feature not a flaw. It is an unusually rare opportunity to see scientific knowledge accumulation in action; flaws, failures, surprises, successes, flights of elation, and all.

## Part 2, The Model

The kind of cognitions that Hawkins would like to implement in machines are best described by the umbrella term General Intelligence. And so, as opposed to A.I. - Artificial Intelligence - it is A.G.I., Artificial General Intelligence that Hawkins is hoping to emulate. A defining feature of general intelligence is flexibility. And flexibility is not a common feature of extant Artificial Intelligence. Current A. I. tends to excel at specific tasks: Winning at games like Chess, Go, and Jeopardy for example. Or carrying-out routine operations like factory floor assembly or relatively well-defined behaviors as seen in self-driving cars. To accomplish these tasks the machines are laboriously trained on large data sets. And after training they can only do that one thing. And if the task requirements of that one thing changes then the machine has to be trained again. That takes time, money, people, and resources. And taking time, spending money, using people and resources are not typically associated with flexibility. But learning on-the-fly, acquiring new behaviors while doing that, not forgetting previously learned behaviors in the process, and doing it all on a lower budget; that could reasonably be thought of as "flexible." Not being confined to fixed behaviors, learning while not forgetting, and doing this in sync with a changing world would, I think, qualify as intelligent in a very general way. Implement this in a machine and voilà, Artificial General Intelligence.

Hawkins characterizes General Intelligence with four attributes: Learning Continuously, Learning via Movement, having Many Models of the world, and Using Reference Frames to store knowledge. The details of the neural architecture supporting

these attributes are beyond the scope of this review. But there is a defining feature common to all four of them that I think is worth understanding in detail. And that feature is Learning Continuously. And learning continuously means acquiring new sequences and remembering them. This feature was mentioned earlier in reference to Hawkins and Blakeslee's assertion that learning sequences of patterns is the most fundamental operation of the neocortex.

There are two settings in which sequence learning takes place. One, when the learner is more passive and the world outside the body is changing, And the other when the learner is intentionally exploring that world. In the tutorial that follows I will look at only the first, when the learner is more passive. But this first kind of sequence learning underwrites the second and is a common and non-optional feature of all four attributes of General Intelligence. So once this basic property is understood then all of the other attributes can be appreciated on their own terms. So with that in mind I would like to turn to illustrating that most important skill of the neocortex, learning and remembering sequences.

### Hierarchical Temporal Memory (HTM)

Figure 1 shows the basic architecture of the neocortex that underwrites sequential learning and memory. The basic graphic that I will use to illustrate the dynamics of sequential learning will be an expanded version of one of the rectangles shown in Figure 1D. Each rectangle represents one of six horizontal layers in the neocortex. And each gray dot represents a neuron within that layer. The model that describes the dynamics of these layers is called Hierarchical Temporal Memory (HTM). A set of capabilities that ensures the acquisition and preservation of temporal sequences. Sequences that are regulated by both hierarchical as well as horizontal reciprocating loops.

Figure 1A shows the outer convoluted surface of the neocortex. Jeff Hawkins points-out that if you could unfurl this it would be about the size of a dinner napkin. And as the text in Figure 1B indicates it is about 2.5 millimeters thick. The small objects in 1B represent neurons. The differences in their size and shape point to the diversity of cell types. And the geometric patterns evident in 1B indicate a characteristic distribution

of vertical columns. These columns transect the six horizontal layers. Figure 1C depicts an expanded view of these columns, referred to as mini-columns in the literature. These are the columns that Vernon Mountcastle noted had such a high degree of uniformity throughout the neocortex. Figure 1D, as mentioned above, is a graphic depiction of the neocortex with the rectangles representing the six layers and the small dots representing neurons. Figure 1E is a drawing of a typical excitatory pyramidal cell, and Figure 1D represents this neuron's logical equivalent in the HTM model.

## How it works

The structure of the neocortical pyramidal neuron is critical to the ability of the brain to learn sequences. Sequences being defined as repeating temporal patterns of neuronal action potentials. There are three zones of influence that independently affect the likelihood of a cell generating an action potential. The three zones of influence are shown in Figure 1E. Two of these are referred to as distal because they are located at some distance from the axon hillock where certain action potentials take place. The third zone is called proximal because of its proximity to this part of the cell body. All three zones consist of dendritic branches that have reception sites for axons from cells in other regions of the brain.

An enlargement of a segment containing some of these reception sites is shown in the box on the right-hand side of Figure 1E. The term apical dendrite is used to refer to dendritic branching that is located at the furthest distance from the cell body. The afferent fibers coming to this region are believed to be bringing feedback to the cell. The basal dendrites, closer to the cell body, are believed to be carrying more local contextual information. And the proximal dendrites are typically thought of as bringing feed-forward inputs to the cell. Prototypical feed-forward inputs would be afferents from the sensory organs of the body; vision, touch, and hearing, for example.

Figure 1F is the HTM, Hierarchical Temporal Memory model of this pyramidal neuron. The circles beneath the horizontal bars feeding into the OR gates represent synaptic locations on the dendritic branches. And each circle represents the terminal end of an axon from a cell in a different part of the brain. It could be from within the same layer or from a more distant region. For the tutorial on sequence learning that we

will look at below these inputs will be horizontal connections from within the same layer as the target cell. Dark dots indicate active, or cell firing, locations on that branch. Therefore, a particular distribution of active sites along a dendritic segment represents a unique spatial pattern of activity into that cell.

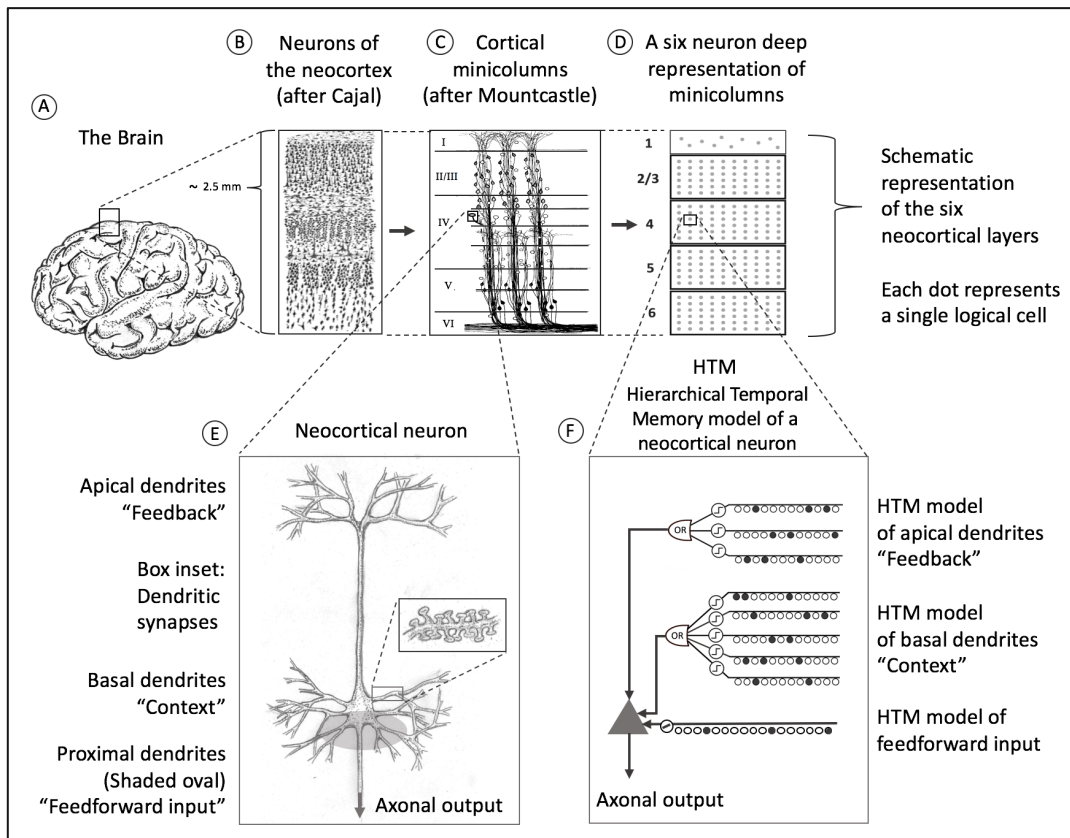


Figure 1. (A) The neocortex. The outermost layer of the brain. (B) Drawing of neocortical layers. Ramon y Cajal. 1911. (C) Neocortical Minicolumns.

Vernon Mountcastle. 1997. (D) Graphic depiction of neocortical slice. Six layers. Dots representing neurons. (E) Drawing of a pyramidal cell. (F) Drawing of the logical structure of a Hierarchical Temporal Memory model of a pyramidal cell.

## Introduction to sequential learning and remembering

Figure 2C represents a cortical macro column in one layer of the neocortex. The vertically aligned dots represent 21 mini columns in that macro column. I place this alongside renderings of neocortical mini columns, 2A, and their HTM counterparts, 2B, just as a reminder of the origin of this graphic that we will be employing. The feature that I want to call attention to is the horizontal connectivity throughout this layer in the brain. Cells in these layers have large-scale horizontal projections to other cells in the layer. If that connectivity were visually represented then the Figure would be a mass of curved lines, so I have drawn only two. Cells (2,1), (4,6), and (13,5) have axons projecting to the dendrites of cell (7,4). And Cells (10,1), (16,5) and (21,1) have axons projecting to the dendrites of cell (15,2). The significance of this resides in how stimulation of cells in widely separated columns can simultaneously affect the firing likelihood of a particular cell in a different column.

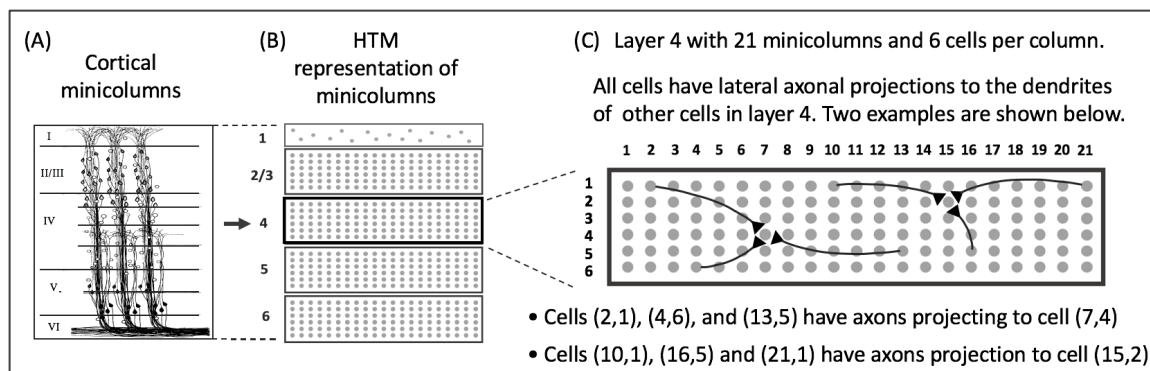


Figure 2. Panel A shows three cortical mini columns transecting six cortical layers. Panel B is the HTM model of these neurons and layers. Panel C is an expansion of cortical layer 4. The dots represent pyramidal cells, the lines represent axons, and the triangles represent synaptic connections onto dendrites.

## Hebbian learning

In this section we will look at how "neurons that fire together wire together." That phrase is credited to the Canadian neuroscientist Donald O. Hebb. The mechanism that enables the firing together is the temporal overlap of a depolarization in the dendrites and a depolarization in the proximal zone of a cell. The cells that are "wired together" are the distant cells that stimulated the dendritic segment and the cell that that segment belongs to.

### Before/During learning

At time T1 consider the pyramidal neuron in Figure 3 to be cell (9,3) in Figure 4A. In this example the axons from cells (3,1), (13,2), and (18,6) converge in close proximity to each other on the same dendritic segment. I have indicated this by darkening the synaptic locations in the dendritic box-enlargement of Figure 3. As mentioned earlier there is massive horizontal connectivity throughout the layer. For simplicity I am just showing one convergence.

At time T2 a feed forward pattern arrives from a sensory organ. It could be a single musical note. This is indicated by the three upwardly pointing arrows labeled Pattern 1. The three arrows index three features of that musical note. They could be the fundamental frequency and two of its harmonics, for example. Each of these features arrives at the proximal zone of all of the cells in their respective mini column, causing all of them to fire. This firing is shown by changing the gray dots to black ones in Figure 4B.

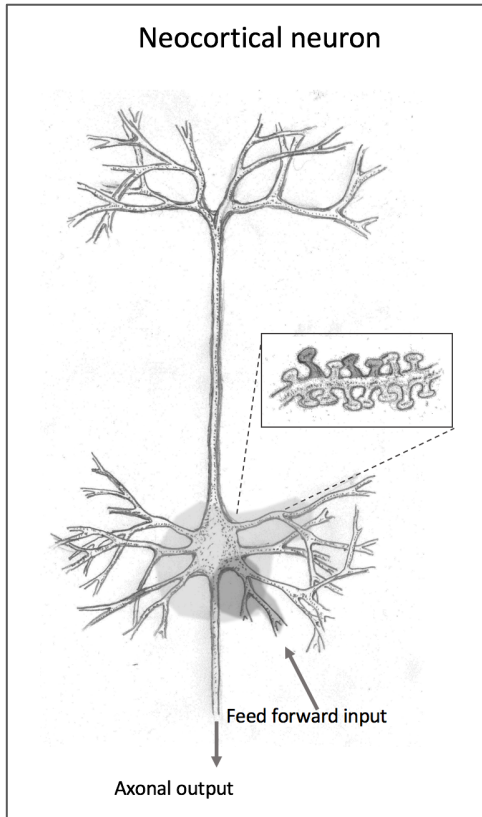
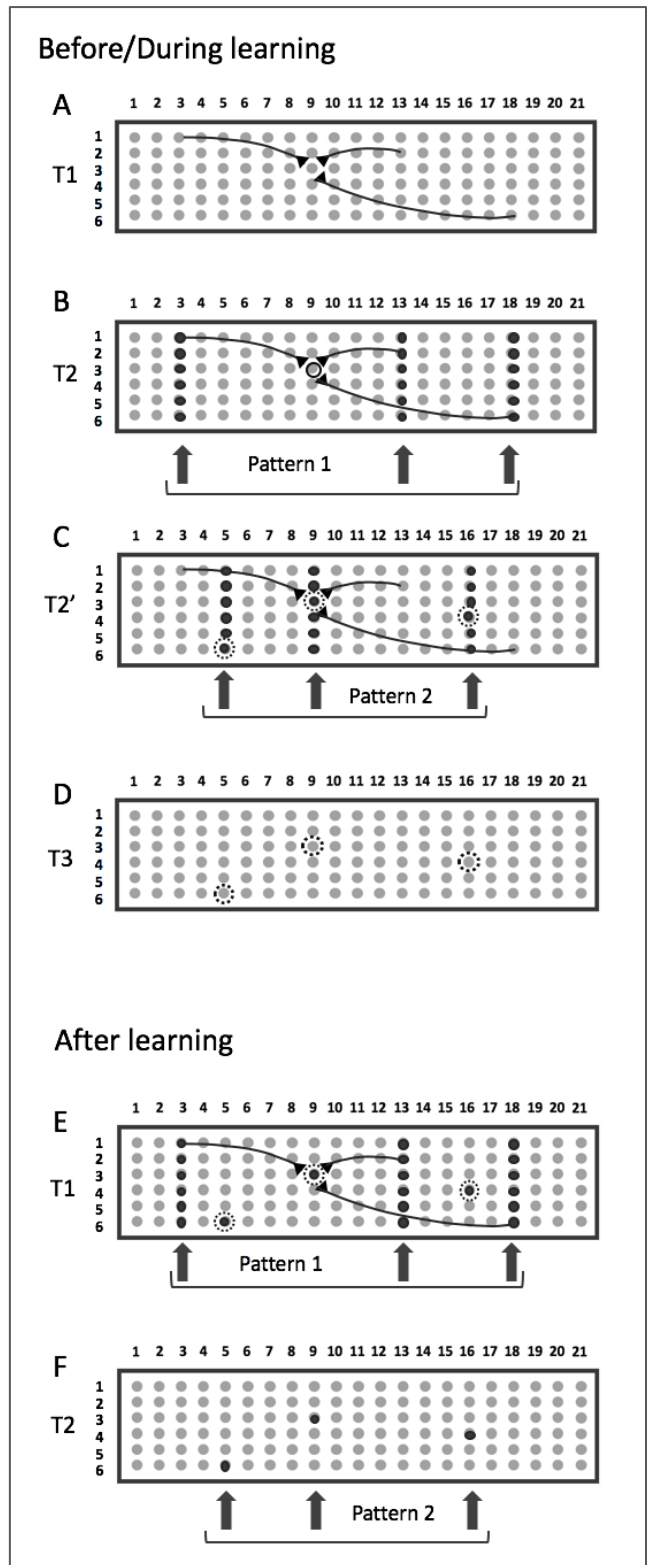


Figure 3. The enlargement box shows synaptic terminals along a dendritic segment. The light gray shading around the cell body indicates a spreading polarization from this dendritic segment. The dark shading represents a spreading depolarization due to a feed forward initiated action potential.

Figure 4. The basics of Hebbian learning. The rectangles represent one layer of the neocortex. The dots represent pyramidal cells. The upwardly pointing arrows indicate feed forward stimulation. Dark dots indicate that a cell is firing. Gray dots are unactivated cells. An outlined gray dot represents a cell with a depolarized dendritic segment. Dark dots surrounded by a dashed circle represent a cell that has active dendritic and axonal depolarizations. Gray dots surrounded by a dashed circle represent a Hebbian cells capable of predicting an upcoming event.



Hawkins and Ahmad (2016) note that learning a new pattern requires about 15 -20 active synapses collocated along a short dendritic segment.. For illustrative purposes in Figure 4, I have shown only three, reflecting the three-feature pattern of columns 3, 13, and 18.

This temporal confluence of these three afferents can be strong enough to cause the membrane of that dendritic segment to depolarize. I have indicated this dendritic depolarization of cell (9, 3) with an outlined grey dot. This local depolarization of the cell membrane can cause the neighboring membrane segment to depolarize which, in turn, can cause the next membrane segment to depolarize, and so on. Thus you can have a membrane depolarization spreading away from the initial dendritic site and expand out across the cell body. I have indicated this with light gray shading in Figure 3. This depolarization can reach all the way to the axon hillock but, typically, it will not cause the cell to fire.

During this time it is possible that a feed forward signal will arrive from a peripheral organ at the proximal zone of the cell. It could be another note in a musical sequence, for example. This is indicated by the upwardly pointing arrows in Figure 4C labeled Pattern 2. I have tagged this event as T2' to indicate that the arrival of Pattern 2 is within the refractory period of the previous dendritic depolarization. That is, the dendritic depolarization is still active. This feed forward Pattern 2 signal can initiate a local depolarization causing all cells in the columns to fire. In Figure 3 I've indicated this depolarization with a darker shade of gray in the proximal zone of the cell. And in Figure 4C by larger dashed circles around a black cell. This event will generate both an action potential travelling away along the axon, and a membrane depolarizing spreading backward along the cell body and out toward the dendrites. This passive spreading depolarization away from the axon hillock and into the cell body toward the dendrites is sometimes referred to as back-propagation, or back-prop for short. What happens next is the Hebbian learning that modifies the active dendritic segment to become a coincidence detector and turn that cell into a predictor of its own state change.



If the spreading depolarization due to the axon potential reaches the recently depolarized dendritic segment while that segment is still active then a chemical process known as synaptic plasticity ensues. This means that the dendritic segment now has a lower depolarization threshold for patterns similar to the one that caused its depolarization in the first place. Specifically, the cell will respond sooner to this input than it did previously. I have indicated this re-biasing in Figure 4D at time T3 with small-dashed outlines around the potentiated cells in mini columns 5, 9, and 16. This ability of cells to fire sooner than they would have before the Hebbian process is critical for the learning and remembering of many different sequences.

### After learning

The axons of Pyramidal cells have local lateral projections onto quickly responding inhibitory cells (Not shown in the diagrams). And these inhibitory cells project to other Pyramidal cells in the mini column. This means that if a particular cell in a mini column fires early it will inhibit all the other cells in that column. Figure 4E shows Pattern 1 arriving again. And as before the features of Pattern 1 activate all the cells in mini columns 3, 13, and 18. I am only showing the afferents to one cell, (9, 3). And, as before, afferent projections stimulate the dendritic regions of target cells in mini columns 5, 9, and 16. But these cells are different now and the depolarization to this stimulation is stronger. The significance of this is made apparent when Pattern 2 arrives.

When Pattern 2 appears as shown in Figure 4F the Hebbian modified cells fire early and immediately inhibit all the other cells in their column. So we see now that, as opposed to the pre-Hebbian event shown in 4C, only one cell in each mini column fires. One way of expressing this is to say that upon the occurrence of Pattern 1 three specific cells in a different region of the layer were prepared to selectively fire upon the occurrence of Pattern 2. They were in an expectant state. Another way of saying this is that three cells in different regions “remembered” that Pattern 2 followed Pattern 1. As I understand it, this is the core process underlying sequential pattern acquisition in the continuous learning model of the neocortex. We can now look at how this works to acquire new sequences.

## Learning Sequences

### Before learning

In Figure 5 the gray upward pointing arrows beneath the layers of cell-representing dots indicate a feed forward input to the layer. In this case a temporal sequence A B C D that could be a melody, for example, of four notes. And, as in earlier examples, the three arrows could represent features of the notes such as frequency formants. Going from left to right in the top row time T1 through T4 is the first time the cortical columns have been exposed to these patterns, so no cells have been put into a predictive state. Therefore these patterns are considered totally ambiguous and all the cells in all of the stimulated columns become active and fire (Dark black dots).

### During learning

As described earlier, cells in the activated (black) columns have previously established connections to the dendrites of cells in other columns (outlined dots). Only one set of lines are shown, projecting to cell (5, 1), but there are also projections from cells in the activated (black) columns to cells (1, 5) and (20, 3) in columns 1 and 20. During the learning phase, a Hebbian connection is established between the feed forward feature-activated (black) cells and their target cells (outlined dots). This connection is brought about by the coincidence of dendritic depolarization due to the three afferents and the back-propagating depolarization due to a feed forward generated action potential.

An example of this learning occurs in the second row when pattern "B" arrives at time T2. Pattern "B" generates action potentials in the columns that contain cells whose dendrites the preceding Pattern "A" has just depolarized. So, for example, Hebbian potentiation takes place at time T2 in cell 1 of column 5. Subsequently this same process repeats between times T2 and T3. The cells that have just been potentiated at time T2 -- (1, 5), (5,1), and (20,3) -- project to the dendrites of other cells in the layer. The dendritic segments of those cells are depolarized. And then, at subsequent time T3, but still within the refractory period of the dendritic depolarization, the arrival of feed forward pattern C initiates an action potential that passively spreads into the dendritic

regions where the two coincident potentials initiate synaptic plasticity. This is now repeated again between times T3 and T4. The process ends at T4 because this is a four note stimulus sequence.

After learning.

The cells that have undergone Hebbian learning now play a new role in the presence of a previously experienced sequence. They function as predictors. In the presence of the beginning of a familiar sequence they not only become active themselves, they also prepare a specific set of other cells in the layer to be ready to respond to a specific upcoming pattern. This is shown in row three. The dots with dashed outlines represent cells that go into a predictive state when the layer receives a previously experienced segment of a stimulus sequence. And in the next time step those expectant cells fire and simultaneously prepare another set of cells to predict the next pattern in the sequence. The fourth row depicts the same temporal event but without showing the preparatory state of the predicting cells.

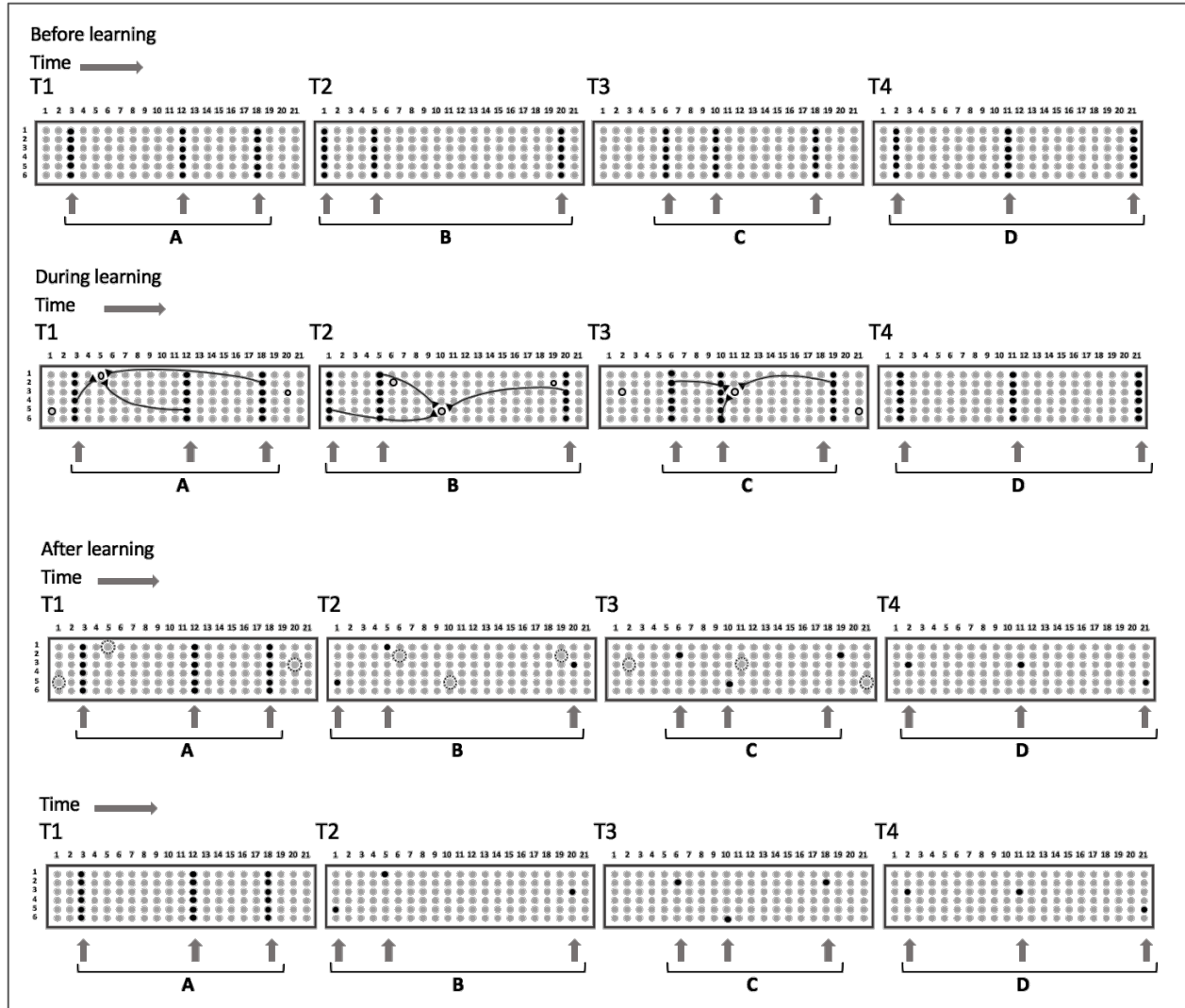


Figure 5. Sequence learning.

The rectangles represent one layer of the neocortex. The upwardly pointing arrows represent three features of a four component temporal stimulus (ABCD). Dark dots indicate that a cell is firing. Gray dots are inactive cells. An outlined gray dot represents a cell undergoing Hebbian potentiation. Gray dots surrounded by a dashed circle represent a Hebbian cell in a predictive state.

## Commentary

The above tutorial was meant to convey the very basics of sequence learning. As a model of how the neocortex works I thought it was worth exploring in detail. And, as we saw earlier, Hawkins and Blakeslee proposed learning and recalling sequences of patterns is the most fundamental operation of the neocortex. To be sure the tutorial was restricted to learning in which the world was doing the changing and the learner was more passive. But, as I mentioned earlier, this elementary process underwrites active learning and is also essential to Hawkins's four attributes of General Intelligence; Learning Continuously, Learning via Movement, acquiring Many Models of the world, and Reference Frames. All of these attributes have prediction as a central part of their behavior. As Hawkins says in Chapter 3, "The brain creates a predictive model. This just means that the brain continuously predicts what its inputs will be. Prediction isn't something that the brain does every now and then; it is an intrinsic property that never stops, and it serves an essential role in learning. When the brain's predictions are verified, that means the brain's model of the world is accurate. A mis-prediction causes you to attend to the error and update the model." (Hawkins, 2021 p. 32).

I would now like to comment on three other model-related issues. One is Sparsity. Sparsity is not discussed in the book. It is however an integral part of the HTM model and is broadly discussed in Numenta's papers and presentations. The other topics concern the establishment of perceptual stability, and efforts to adjust the model to take an active agent into account. I will begin with Sparsity.

## Sparsity

A stranger comes up to you on the street and says, "There are too many tutus to count." You might be personally annoyed, curiously puzzled, or lightly amused. But one thing you would not be is lexically confused. But why not? Acoustically all those "to"s are the same stimulus. Could this strange term "sparsity" have anything to do with your lack of confusion? Lets ask Jeff.

Sparsity is a fundamental part of HTM theory. The term of art in the model is SDR, Sparse Distributed Representation. It is important because the brain appears to work on this principle and Numenta's goal is to re-engineer the brain. As Ahmad and Hawkins say: "Cortical representations are such that only a small percentage of neurons are very active while the rest remain relatively inactive, and hence the activity is considered sparse. (Ahmad & Hawkins, 2015. p.1).

Sparsity means just what it says, "less." Although, in this case, less is more. For example sparse data structures have a very high representational capacity, they are good at recognizing patterns in the presence of noise, they are highly tolerant of random deletions, and they are very efficient. Numenta has excellent papers and videos demonstrating these features. See examples in the reference section. Here I only want to convey what sparsity is, and give a few examples of how it works.

In conventional computers the opposite of sparse is dense. Dense means that all of the available bits are used to represent a feature or a category. An example of this is the American Standard Code for Information Interchange, or ASCII code. ASCII codes represent alphanumeric symbols in computers and other common telecommunication devises.

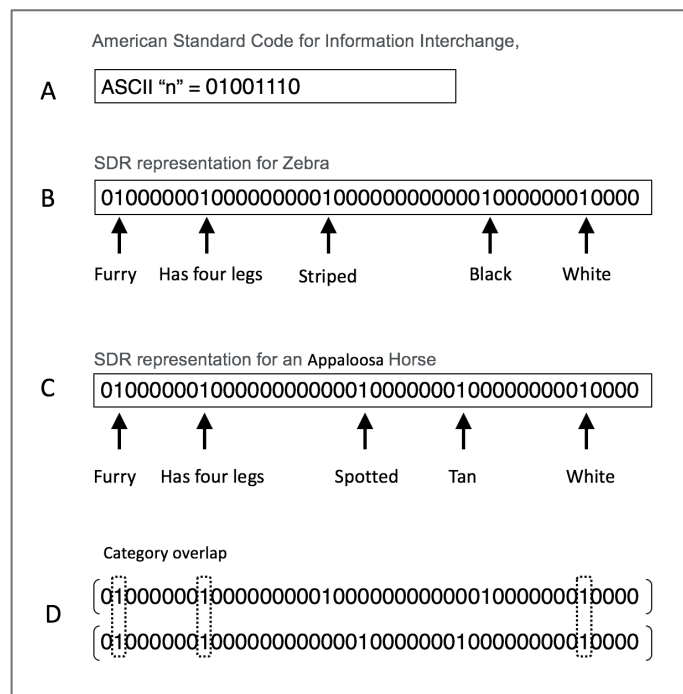


Figure 6. Row (A) - ASCII code for the letter "n". Row (B) - Sparse Distributed Representation for a zebra. Row (C) - Sparse Distributed Representation for an Appaloosa horse. Row (D) Overlap representation of the SDRs for zebra and horse.

Figure 6A shows the ASCII code for the letter "n." By contrast figure 6B is an SDR representing a zebra. (Hole & Ahmad 2021, adapted from figure 5). A major difference between the two, besides length, is that the individual bits in the ASCII code have no intrinsic meaning. The meaning of an ASCII code is in the assignment of the entire 8-bit pattern to a given alphanumeric. Whereas each bit the the SDR represents some feature. For example, each of the bits indicated by a "1" in figure 6B represents a feature that is common to a zebra.

All of the bits in the ASCII code are required if the string it is to function as intended. If just one bit is missing or changed the meaning of the code is lost. The SDR however can survive the loss of many bits and there will still be some degree of category representation preserved. This is just one example of the robustness of SDRs against noise or random deletions. If the bits in the SDR are regarded as active or inactive neurons, then this representation is considered to be *distributed* because the information is encoded not in just a single neuron but across a set of active neurons.

A simple example of how an HTM algorithm would use such representations would be the comparison of an SDR acquired from scanning an image of an animal, for example, to a standard SDR of a zebra. This is done with what is called an overlap function. In this case the overlap of figure 6B with figure 6C, and shown in figure 6D is 3. Typically there is a pre-assigned threshold for a match. If that threshold was 3 then this would be a match. Clearly, in this case, that threshold is too low, since a horse is not a zebra. In HTM models the SDRs are typically much longer, from 1024 to 65,536 bits for example. And the sparsity set anywhere between .05% to 2.0%. And the number of features can vary also.

It might appear, at first glance, that using such a small fraction of your available bits would severely constrain the number of unique SDR encodings. But that is not the case. There are 44 bits available for categorization in this zebra example. So even with a small vector like this, and a sparsity number of, say 5 bits, you get 1,086,008 possible encodings. That's not a small number. But, in any case, this is for a relatively small demonstration vector. Vector size in realistic models are well over 1000. So, even using the low end of 1024 bits and a sparsity of 10 bits you are looking at an encoding space of a three-digit number with 21 decimal places to the right of it. You're pretty much good to go.

## Context

Another thing that SDRs are particularly good at is identifying the same features as having different meanings when they occur in different contexts. Figure 7 shows the response of the mini columns of an input layer to the musical note "A." On the left, under the heading

"During learning," we see the context in which the note "A" first occurred. There are four context examples: "A" was not preceded by anything, "A" was preceded by the note "B," "A" was preceded by the note "C," and "A" was preceded by the note "D." On the right we see a two-time-step response of the layer to the note "A" after Hebbian learning has occurred.

In the top row, where no note appears at time  $t=1$ , the layer treats the note "A" at time  $t=2$  as a completely novel event and all of the cells in the the three-feature column become active. In the second row a "B" occurs at time  $t=1$ . Because of the previous Hebbian learning process there are cells in the features columns of note "B" that put other cells in the layer into a predictive state. The cells in the predictive state are indicated by an open circle symbol. Now, at time  $t=2$ , when note "A" arrives, these predictive-state cells fire first and inhibit all others. This is indicated by a black dot in columns 3, 7, and 18. Note, this is still recognized as an "A." The single cell firings are in the same columns as the unbiased "A" in the first row, that is 3, 7, and 18. What is unique at time  $t=2$  in the second row is that these particular cell firings carry a "memory" of the recent history of this "A."

A similar pattern occurs in the third row example when a "C" note arrives. This "C" note had, during learning, formed a Hibbian connection to certain other cells in the layer. So, at time  $t=1$ , it activates the dendrites of those cells and those cells go into a predictive state as indicated by the open circles of cells (3, 6), (7, 5) and (18, 1). Now, when note "A" arrives at time  $t=2$ , only cells (3, 6), (7, 5) and (18, 1) fire. And, as before, the system recognizes this as an "A" but an "A" with a unique history. The last example in the fourth row reflects this same process. The arrival of note "D" puts cells (3, 1), (7,2), and (18,5) into a predictive state at time  $t=1$ . Then, with the arrival of an "A" at time  $t=2$ , only those cells become active, thus reflecting a unique Hibbian history.



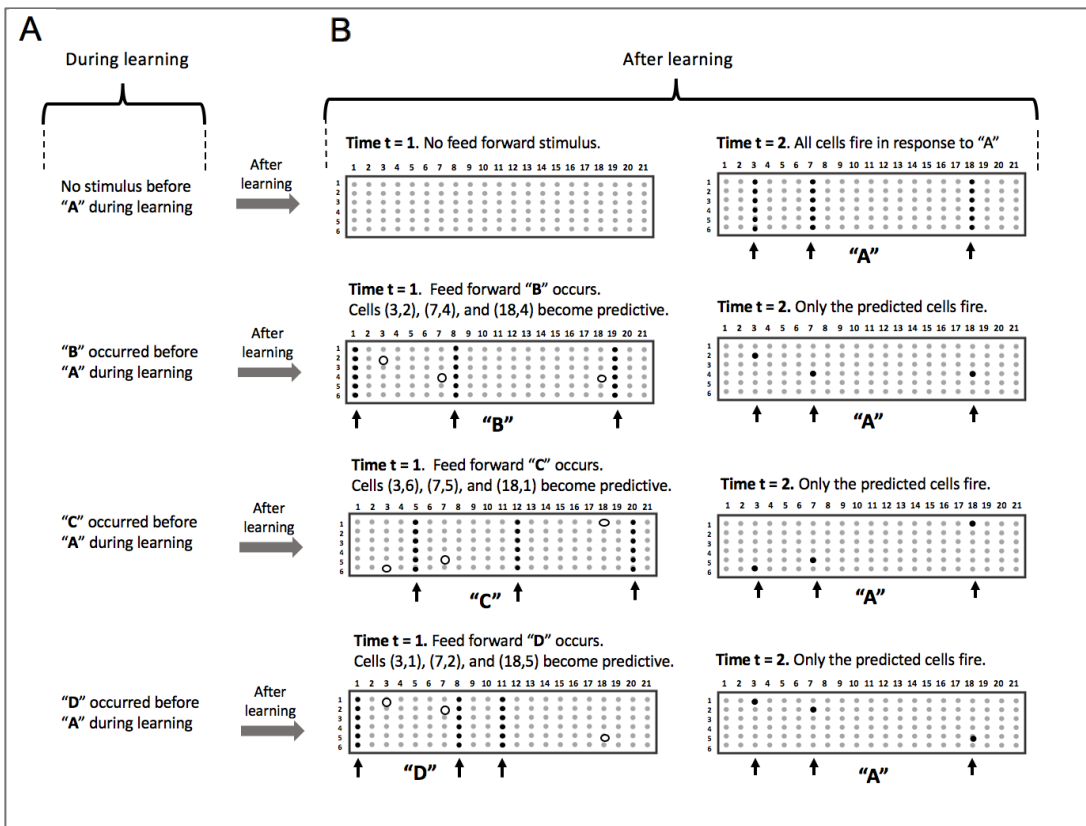


Figure 7. Context. Sparsity's role in establishing the uniqueness of a feature in a particular context. (A) During learning. This column shows four examples of a context immediately prior to learning the stimulus "A." They are, no stimulus, or a "B" or a "C" or a "D." (B) After learning. A two time-step demonstration of unique learning. In the t=1 phase the context stimulus activates all the neurons in their feature minicolumns. But due to prior Hebbian learning the "B," "C," or "D" prime different cells to enter into a predictive state. Therefore, when the "A" stimulus arrives it will activate different cells in the "A" features mini columns. This is a signature that each of these "A" has a different meaning.

What about capacity? Once again, this relatively simple sparse system of 21 mini columns with 6 cells per column has a very large classification capacity. That is 21 active columns of six cells per column gives  $6^{21}$ . That is, six raised to the power of 21 possible independent sequences. This is quite a large number. And, once again, it is just for a small demonstration example.

Figure 8 is an example of how networks like the ones we have been looking at need to function in the "real world." That is they need to simultaneously hold a lifetime of different sparse

sequential representations. I won't illustrate a lifetime of sequences. But figure 8, adapted from figure 1 of Cui et al, is meant to show how a system that can represent the same input in very many different ways can learn very long and complex sequences and never get confused. (Cui et al., 2016)

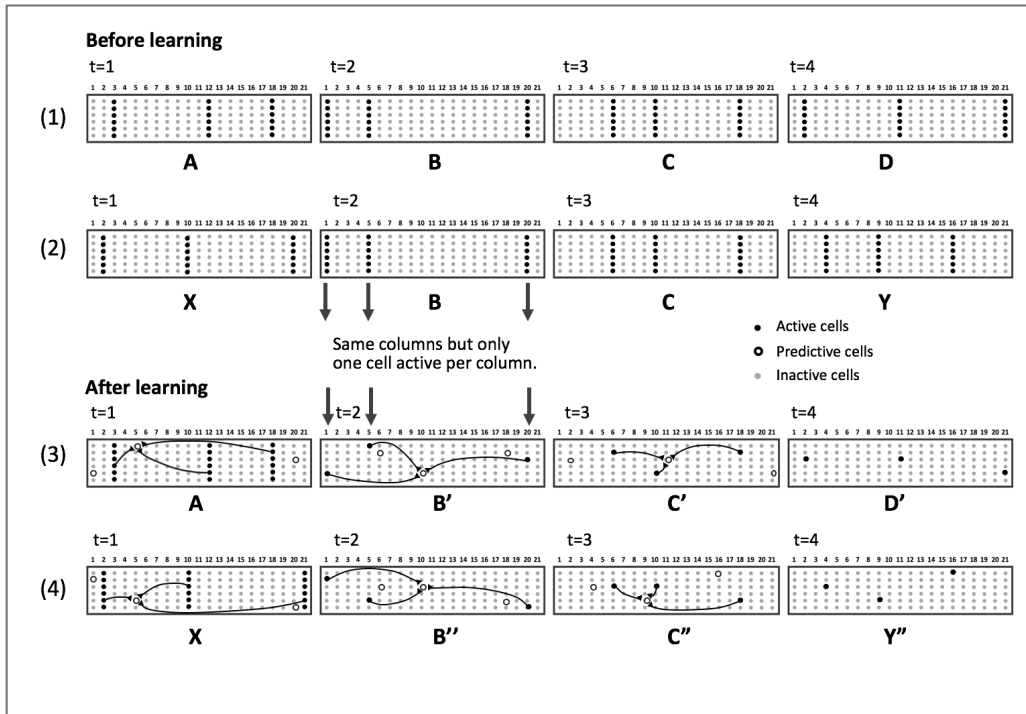


Figure 8. This figure illustrates how a single cortical column can simultaneously hold many different sequences. In particular how different sequences that share the same internal segments can, nevertheless, not get confused, and still correctly produce different patterns at the end. See text.

Figure 8 illustrates the learning of what Cui et al. call high-order sequences with shared subsequences, ABCD vs. XBCY, for example. Figures 8(1) and 8(2) show the learning phases that we have seen before where all of the three-feature inputs are novel and thus all the cells in the feature columns fire. I am not showing all the upward arrows in this graphic. But the labels ABCD and XBCY could be musical notes as before, or any other stimuli that has three features. The reason that this system will not be confused by similar imbedded subsystems is due to the fact that once learning has taken place the model maintains two simultaneous representations. One is at the mini-column level which represents the current feed forward input. And the other is at the dendrites of individual cells representing the context of that input. So in figure 8(3) at time

t=2 when a "B" that was predicted arrives, it activates only those cells in each of the "B" feature columns that were predicted by the previous stimulus "A." This "B" is labeled B' in the graphic. Had this sequence been that shown in figure 8(4) where the opening stimulus was an "X," then the arrival of "B" at time t=2 would have activated three *different* cells in the "B" features mini columns. This is labeled B" in figure 8(4).

At this point the two different cascades of correctly predicted sequences have baked into them a residual of their unique origins; an "A" in one case and a "X" in the other. Thus B' will predict a C' and a B" will predict a C". And as Cui et al. summarize it, "because different cells respond to "C" in the two sequences (C' and C"), they can invoke the correct high-order prediction of either D or Y." (Cui et al., 2016, p. 2480).

In this example it was a single event, an "A" or an "X" that uniquely biased all the subsequent activations. But in the real world that "A" or that "X" is more likely to be an arbitrarily long sequence of president-setting precursors. Because of these Sparse Distributed Representations the HTM models can learn very long-range temporal dependencies. As an example Hawkins and Ahmad calculate the number of possible representations for a layer of 40 active columns with 32 cells in each column and one active cell per column. "Then there are  $32^{40}$  possible representations of each input pattern, a practically unlimited number." (Hawkins and Ahmad, 2016, p. 9). I don't know the magnitude of a "practically unlimited number," but my guess is that it is pretty big. And perhaps this is one of the reasons that Hawkins said at talk he gave at Berkeley: "If you're going to forget everything else I talk about today and tomorrow and you want to remember one thing from my talk, we can remember this. And I'll state this right very clear, I'm certain of this. That all intelligent machines, biological or otherwise, are going to be based on sparse representations."<sup>1</sup> So, tell us what you really think, Mr. Hawkins. Why hold back?

In his book Hawkins recounts the story of an invited talk he gave at Intel in 1992. In his talk he proposed that the future of personal computing would be dominated by devices small enough to fit into your pocket. Everyone knows that pit-in-the-stomach feeling of having to "go on" right after a talk given by a brilliant, well-liked, and charismatic speaker. What is less well know is what it is like having to end your talk before an audience who knows that lunch will not be served until the Q and A session is over. Lunch delayed can bring out the worst in a questioner. Hawkins remembers one query: "What are people going to use these handheld

---

<sup>1</sup> Berkeley talk. <https://www.youtube.com/watch?v=qZM9JREjnp4>. Location 33:01.

computers for?" I have the impression that the answer was lost in the clatter of serving trays, rising chatter, scooting chairs, and waiter talk.

After speaking Hawkins was seated at the lunch table with Dr. Gordon Moore. That's the Gordon Moore of Moore's law. So there you are. Napkin on your lap, fork and knife in the correct hands, music playing over the P.A. Why not ask one of the most revered tech gurus in the world what he thought of your talk. The details of that moment can be found on page 116 of the book. But the bottom line Hawkins says was, "Starting that day, I avoided giving 'visionary' talks about the future of computing and instead did as much as possible to make that future happen." There are two things to note about this. He did just the right amount, and then some, to make that future happen. And, as we see from his "If you're going to forget everything else I talk about today" prophesy, he got over that shyness about visionary talks. From whence comes this confidence? I'm going to cherry-pick my way through some of my favorite Hawkins features and conclude that he does indeed have something to be confident about.

It's hard to be sure where confidence comes from, but try Donald O. Hebb, Karl Lashley, and Vernon Mountcastle. I think Hawkins turned to the right people to make sense of the last half-century or so of neurological discoveries. For starters, inspired by Mountcastle, he realized that the self-similarity of the entire neocortex was probably that way for a reason. Perhaps regions that are wired in similar ways fire in similar ways. Enter Karl Lashley and serially ordered behavior. The world comes to us in dynamic non-stationary temporally ordered sequences. Pardon my jargon, but the point is that creatures who move in the world not only have to keep up with current affairs, they need to be ahead of the game. So partner Mountcastle and Lashley together with current understandings of the brain as an organ of prediction and you're almost there. What makes it all run? Yes, Donald O. Hebb. Credit Hawkins for noticing that the the elemental gating unit is not one of those "point neurons" in a node, hidden in a layer, adjusting weights on identical inputs, at the bottom of an algorithmic tree. It is, rather, a three input cell having a feed-forward connection on one of these inputs and thousands of appearing and disappearing synapses on the other two. So, if you go with the wiring, take note of the timing, and in good faith ignore the unknowns for the moment, you plausible end up with a Hierarchical Temporal Memory complete with Sparsely Distributed Representations. And along with that enough confidence-building to in good faith say to an audience that if they're going to forget everything else you have to tell them today, remember this: "That all intelligent machines, biological or otherwise, are going to be based on sparse representations." So, stay tuned. But,

in the short term you may now know why you may have been annoyed by the tactlessness, thoughtlessness, and temerity of that stranger, but not too confused by his tos, toos, and tutus.<sup>2</sup>

---

<sup>2</sup> Talk given on University of California Television (UCTV). 2012.  
<https://www.youtube.com/watch?v=qZM9JREjnp4>  
Location 43:30.

## Neurological correlates of perceptual stability

In the fourth row of figure 5 we see ever changing patterns of the sparse encoding of feed forward input. You might think that these neurological activities would be the correlates of perceptual experience. But our perceptions of the world are stable experiences, not ever shifting patterns. So you might ask, does Hawkins's model have a mechanism for creating perceptual stability out of these shifting patterns? The answer is Yes. And it's not complicated, he says. (Hawkins, 2021, p. 102). The brain takes a vote. Or, rather, the neocortex takes a vote. Or, better yet, something like 150,000 cortical columns settle on the best guess about a shifting set of SDRs regarding what's happening in the world. Now one person's "not complicated" can be another person's "not comprehensible." Nevertheless with the model of learning and predicting that we have looked at up to now I think Hawkins is more-or-less right about this.

Figure 9A is a graphic representation of the "voters." It represents a section of the neocortex, the tip of a finger for example. These are Mountcastle's cortical columns. The columns mentioned earlier that from the front of the brain to the back of the brain, from side to side, and from the top to what there is of an underside, have more structural commonalities than they do differences. Which suggests that they are all also performing the same or a very similar algorithm, with the differences arising from the diversity of structures they are connected to. This is part of the reason that the voting mechanism is not such a complicated one. It has been standardized, so to speak.

It should be noted that these columns shown in figure 9 are more conceptual conveniences than they are anatomical realities. They are basically defined by several hundred mini columns of the kind we have been looking at in the examples above. Each mini column responds to a feature of the stimulus, and the macro columns are thus defined by a set of features detectors that have some functional or sensorial similarities. In the tactile domain this would be the small patch of skin in contact with the contours of an object being categorized. Figure 9B is a depiction of these mini columns along with an HTM representation.

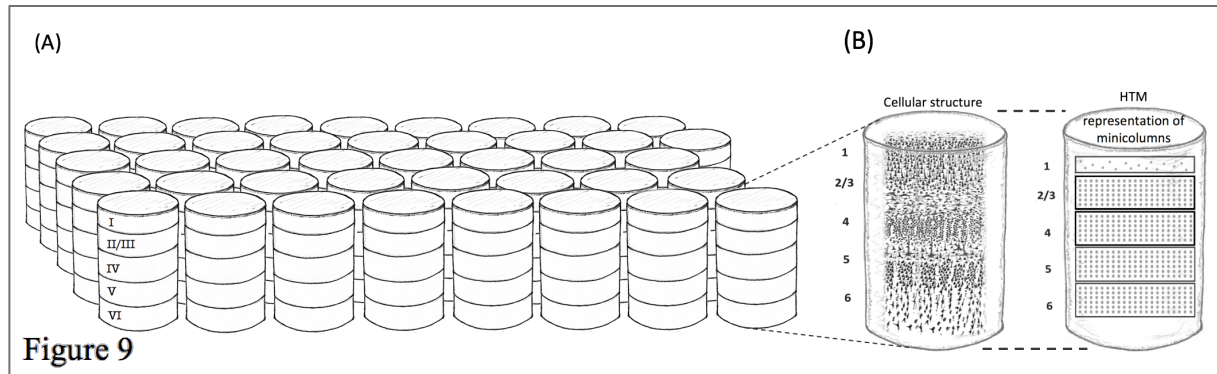


Figure 9. (A) Idealized depiction of the stacked columnar layers of the neocortex. (B) Drawing of a vertical slice of the six cortical layers and a drawing of an HTM model of those six neocortical layers.

## Perceptual stability

Hawkins describes achieving perceptual stability in the face of continuous sensory flux in chapter 7. He precedes his discussion with an intuition-priming section on people trying to find out what city they have been dropped into by using feature-bearing maps and a communication system connecting them with people in similar straits located in other cities. It is a good analogy that I encourage the reader to spend time with. But for here I'll just stay with the HTM representations that we have been using up to now.

Hawkins's description of voting and perceptual stability in his book does not include graphics that could depict exactly what layers are doing precisely what. Since I will be using the HTM model in my discussion of perceptual stability I want to try and be as accurate as possible. To do that I will draw upon a video of a talk that Hawkins gave to the Simons Institute for Theory and Computing in 2018 and that is available for viewing through Numenta's website.<sup>3</sup>

The process of perceptual stability is carried-out by a two-layer input-output circuit that is commonly found in all cortical regions. My figure 10B illustrates this as a section of the HTM cortical column. This figure was inspired by the graphics that

<sup>3</sup> <https://www.youtube.com/watch?v=zVGQeFFjhEk&t=41s>

Hawkins used in his Simons Institute talk. In particular, starting at time location 21:54 of the talk, where he illustrates a two time-step sequence of forming a prediction. His graphics are the same as those I used in figure 5 for example. At location 21:58 Hawkins explicitly calls-out this layer as layer 4. And he describes the construction of predictions exactly as we seen in figure 5 of this review. Then at location 23:38 he shows layer 4 feeding into layer 2/3. And, lastly at location 26:11, this input layer is shown with a broad horizontal double headed arrow inside of it. I take this to be indicating that, at least across the width of a column, but not extending beyond the column, layer 4 has the lateral projections necessary to facilitate sequence learning and prediction.

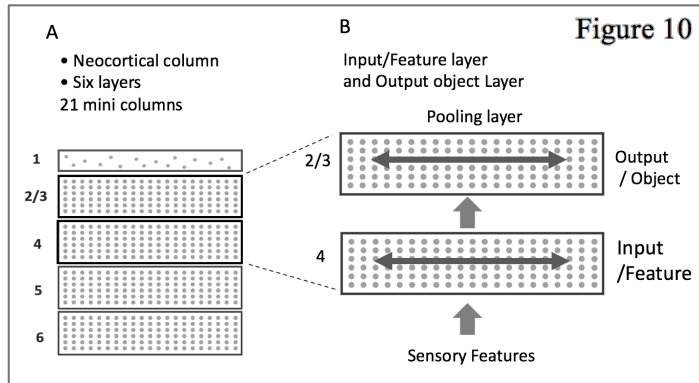


Figure 10. Hierarchical Temporal Memory model of the six neocortical layers. (B) Drawing of cortical layers 2/3 and 4. Vertical arrows representing stimulus information flow direction. Horizontal arrows representing lateral connectivity of within-layer neurons.

Figure 11. Four temporal increments representing the Hierarchical Temporal Memory categorization of a cup out of three possible candidates.

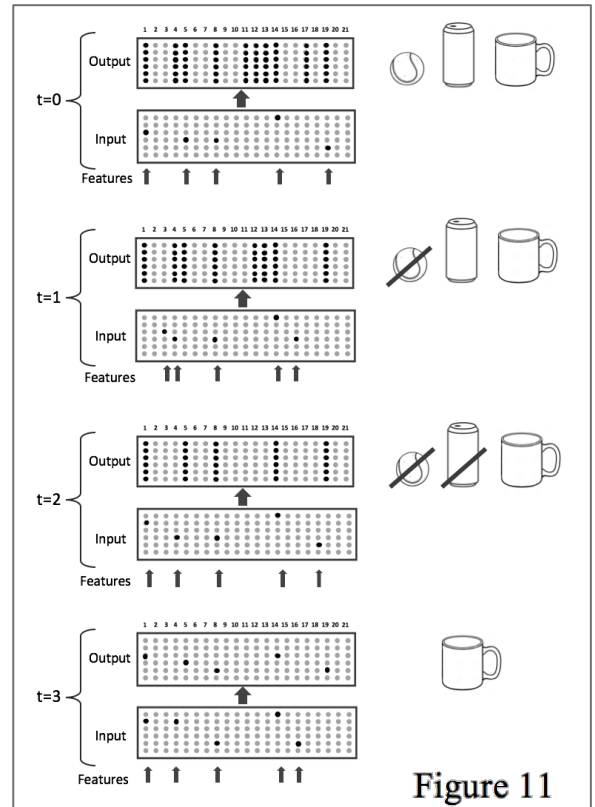


Figure 11



So figure 10 B depicts an architecture in which an input/feature layer sequences Sparse Data Representations into a pooling output/object layer. Since this process is occurring within one cortical column it is not technically a voting process. This is because voting is something that takes place between columns not within them. Recall that the vertical mini columns of gray dots represent sensory features each of which has a defined receptive field. And therefore the output of the features layer would be a Sparse Data Representation carrying a winner-take-all kind of summary of the identity of that small sensory region, a patch of skin on the finger, for example. Figure 12 shows a different situation in which voting does take place. This is a case in which three sensory summaries are independently passed to the pooling layer which, in turn, forms unions where the dominant feature-overlap is established as a stable representation of the sensed object. We will look at that in a moment. For now, returning to the single cortical column, figure 11 illustrates the temporal process in which a single cortical column makes its contribution to the voting pool. It also demonstrates something that people seem to find confusing about the HTM theory. And that is how one single column can learn a complete model of the world. Or more practically an object in that world.

At time  $t=0$  a small patch of the finger makes contact with a surface the character of which is common to all three objects depicted on the right hand-hand side. The SDR passed up to the Output/Object layer is maximally ambiguous and so all the columns in that layer light-up. Indicated by black dots in the columns. Next, at time  $t=1$  the finger moves to another location and, for example, encounters a rim. Since baseballs do not have a rim a baseball is eliminated as a candidate category. I have shown a reduction in the number of active columns in the output layer as an indicator that uncertainty is being reduced. The finger moves again at time  $t=2$ , and it encounters the connecting arc of a handle with a cylinder. Since soda cans do not have handles a soda can is eliminated. And the Output/Object layer simplifies further. Finally, at time  $t=3$ , the finger moves to a new location which is completely compatible with a coffee cup. I've indicated the stability of the classification by darkening only one cell in each of the columns that became active at time  $t=2$ .

This could be thought of as the stabilization of the broad class of "a cup" to a narrow index of this is "my cup."

I have assumed, as I suppose Hawkins does also, that the algorithms operating in all of these layers are the same as, or similar to, the ones we have looked at earlier in the formation of sequences and predictions. That means that this one cortical column will remember the encounters that it had at all locations on the cup. Said another way, that particular cortical column has, indeed, formed a complete model of this little part of the world. Just like Hawkins says it does.

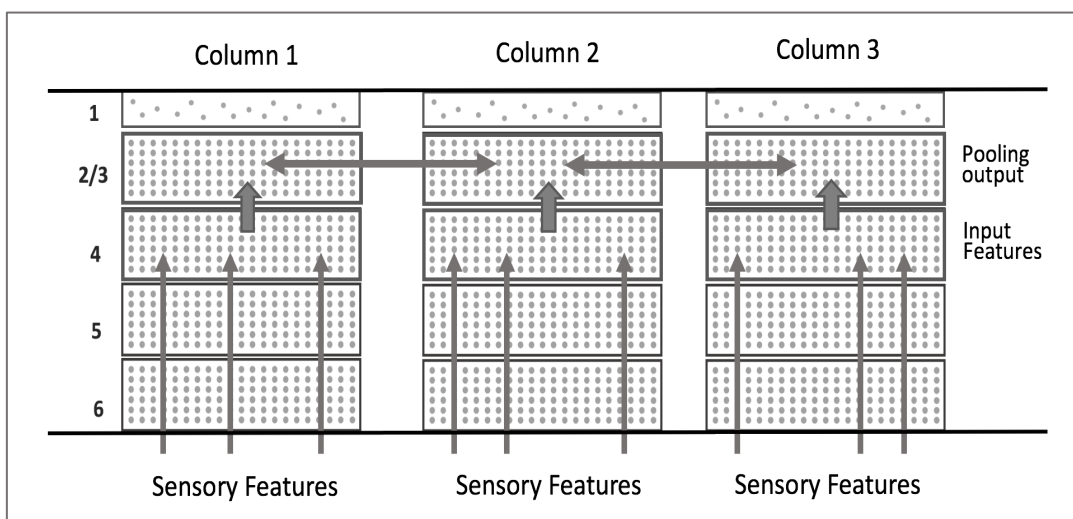


Figure 12. HTM model of how voting takes place between three neocortical columns. Thin vertical arrows represent a feed forward stimulus. Wide vertical arrows represent "winner-take-all" SDRs from layer 4 into layer 2/3. Horizontal arrows represent lateral connectivity between macro columns. Pooling takes place within the columns. Voting takes place between columns.

As mentioned earlier figure 12 illustrates the condition of multiple sensors encountering an object at the same time. In this case what Hawkins calls voting occurs. Here the thin upwardly pointing arrows might indicate features of the object sensed independently by three different fingers. Summaries of these features are passed on to Pooling layers which have multiple lateral connections to neighboring columns as well as other distant parts of the brain. Some of these distant areas might seem unlikely. Cells in some layers, Hawkins notes, "might send their axons from one side of the brain

to the other, for example, between the areas representing the left and right hands. Or they might send their axons from V1, the primary visual region, to A1, the primary auditory region." (Hawkins, 2021, p.102). This makes sense from a model-theoretic point of view. And that's because object stability must occur in all sensory modalities. A cup that has been established as "my cup" by the fingers must also be confirmed as "my cup" by the eyes.

The goal of the brain, Hawkins comments, is to reach consensus about what is out there. He uses the famous example of the shifting faces or vases image to illustrate this. The image can appear as either a vase or two faces, he notes, the brain "does not permit two objects to be active simultaneously—so it picks one possibility over the other. You can perceive faces or a vase, but not both at the same time." (Hawkins, 2021, p. 104).

Figure 13 is just an homage to Jeff Hawkins's cup-whispering skills as they might have been in play on that late February day in 2016. Highly idealized, the illustration shows a three time-step disambiguation process. The temporal column on the right depicts a continuously changing input pattern while the pooling/output stage achieves categorical stability.

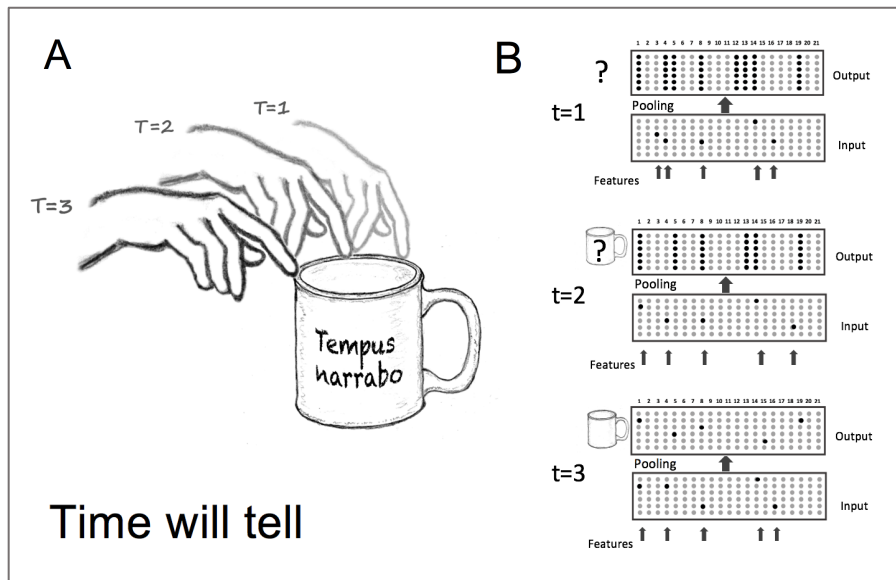


Figure 13. A cartoon version of a three-step disambiguation of a cup. (A) the finger takes three samples. (B) HTM rendering of the disambiguation process.

Figure 14 is an adaptation of figure 1A in Hawkins, Ahmad, and Cui's 2017 paper, *A Theory of How Columns in the Neocortex Enable Learning the Structure of the World*. This was the first Numenta paper published after Hawkins's February 2016 realization that a neurological correlate of a body-independent framework was needed to complete his theory. What is new in figure 14 compared to figure 12, for example, is the addition of an extra arrow feeding into the input layer along with the sensory stimulation. As Hawkins commented in his description of that moment. "This observation means there must be neurons in the neocortex that represent the location of my finger in a reference frame that is attached to the cup. The movement-related signal we had been searching for, the signal we needed to predict the next input, was 'location on the object.'" (Hawkins, 2021. pp. 48-49). The key phrase there is "movement related signal." As Hawkins emphasizes many times, we learn the world through movement, not by passive experiences. That is why solving the "melody problem" was not enough. The agent has to intentionally explore the world to learn about it. All of the examples we have been looking at, sequence learning, prediction formation, and even perceptual stabilization, were "floating" so to speak. That is, there was nothing intrinsic to the algorithm that would inform the agent that these experiences were not just part of the body. Hawkins draws our attention to this issue in the opening to Chapter 5 where he asks, "Why do we not perceive objects as being in the eye?" And, "If the chairs, desks, and courthouse are imaged next to each other on my retina, how is it that I perceive them to be at different distances and different locations?" The same holds for sound. We perceive the sound of an object at some distance and some orientation away from us and not at our ear where the "sound" actually is. (Hawkins, 2021. p. 57.)

This brings us back to Hawkins's Archimedes moment between himself and his coffee cup. The reasoning is sound. There has to be some extra something in the neurological flurry that allows for a "last minute" decision to move our finger right or left, say, and still prepare the correct sensation for the location where the tip of that finger will alight. To say that this is a non-trivial problem is an understatement. And I would also like to say that the solution is beyond my pay-grade. I do have thoughts on the matter. But why embarrass yourself if you don't have to. I will say, however, that I have

been spending a little more time with my coffee cup than before. You know ... you never know.

I can say that this is an ongoing issue at Numenta. How to account for the actions of the perceiving agent and how to establish independent reference frames for the objects and events that are perceived. When you read the book, particularly chapter 5, you will see that Hawkins evokes the structure of the phylogenetically older hippocampal complex as part of his modeling of the neocortex. This is not unreasonable in my opinion. In that complex there are cells that have been given names like "Place Cells," "Grid Cells," "Head Orientation Cells," and "Vector Cells." You can probably guess from their names what these cells have been identified to do. Way back in time, well after our old friend J.T., but well before the advent of a cerebral cortex, these cells helped animals orient and move their whole bodies in the local surround. The suggestion here is that, over evolutionary time, parts of the hippocampal complex was duplicated, refolded, and relocated. This was done as creatures developed more complex appendages; appendages that could utilize previously established spatial capacities for more fine-grained movements and explorations.

My guess is that this is going to work out. But they are not popping the champagne at Numneta yet. I do suggest reading the 2017 and 2019 papers and then viewing the videos made available online. Numenta also encourages participation in their meetings so, don't be shy. You, and your coffee cup, will be welcome.

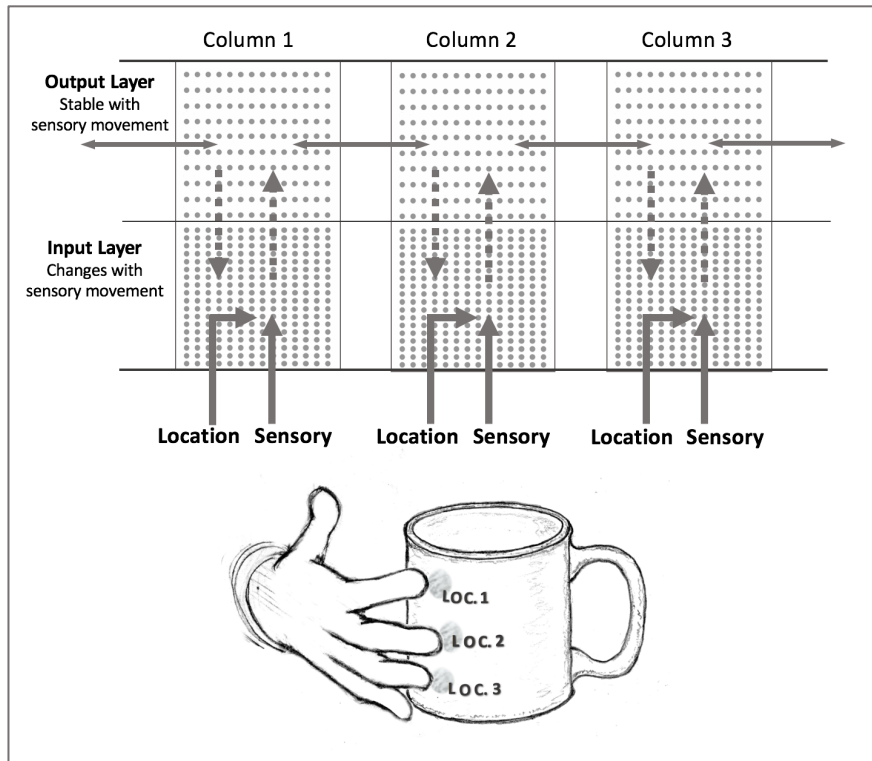


Figure 14. Diagrammatic representation of the information flow during the HTM process of perceptual stabilization of three independent samples of the features of a cup. Adapted from figure 1A in Hawkins, Ahmad, and Cui, 2017.

## References

[https://www.ted.com/talks/daniel\\_wolpert\\_the\\_real\\_reason\\_for\\_brains?language=en](https://www.ted.com/talks/daniel_wolpert_the_real_reason_for_brains?language=en)

The real reason for brains TEDGlobal. July 2011

Crick, F. H. (1979). Thinking about the brain. *Scientific American* 241, 219–232.

Mountcastle, Vernon. “An Organizing Principle for Cerebral Function: The Unit Model and the Distributed System.” In *The Mindful Brain*, edited by Gerald M. Edelman and Vernon B. Mountcastle, 7–50. Cambridge, MA: MIT Press, 1978.

Hawkins, Jeff, Marcus Lewis, Mirko Klukas, Scott Purdy, and Subutai Ahmad. “A Framework for Intelligence and Cortical Function Based on Grid Cells in the Neocortex.” *Frontiers in Neural Circuits* 12 (January 2019): 121.

Hawkins, Jeff, and Subutai Ahmad. “Why Neurons Have Thousands of Synapses, a Theory of Sequence Memory in Neocortex.” *Frontiers in Neural Circuits* 10, no. 23 (March 2016): 1–13.

Hawkins, Jeff, Subutai Ahmad, and Yuwei Cui. “A Theory of How Columns in the Neocortex Enable Learning the Structure of the World.” *Frontiers in Neural Circuits* 11 (October 2017): 81.

Lashley, K. (1951). “The problem of serial order in behavior,” in *Cerebral Mechanisms in Behavior*, ed L. Jeffress (New York, NY: Wiley), 112–131.

Hebb, D.O.: *The Organization of Behavior*, Wiley: New York; 1949

Ahmad S, Hawkins J (2015) Properties of sparse distributed representations and their application to Hierarchical Temporal Memory. [arXiv: 1503. 07469](https://arxiv.org/abs/1503.07469)

Hole, Kj, Ahmad, S (2021) A thousand brains: toward biologically constrained AI  
SN Applied Sciences 3:743 | <https://doi.org/10.1007/s42452-021-04715-0>

Cui Y, Ahmad S, Hawkins J (2016) Continuous online sequence learning with an unsupervised neural network model. *Neural Computation* 28:2474–2504.

Hawkins J, Ahmad S (2016) Why neurons have thousands of synapses, a theory of sequence memory in neocortex.  
*Front Neural Circuits* 10:23