Corey Gaber

An Empirical Attack on the Traditional A Priori

An Empirical Attack on the Traditional A Priori

How the Organ of Cognition Works

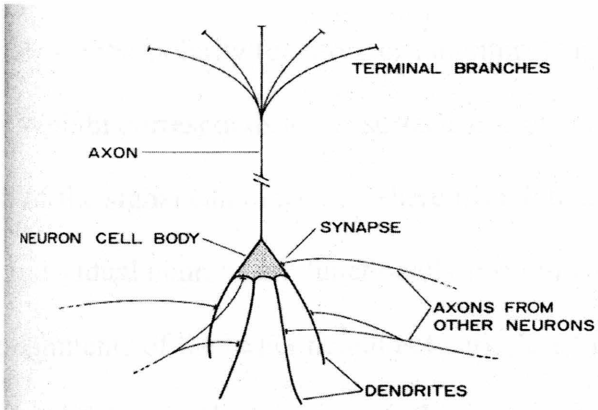


Figure 9.1
A schematic neuron

The brain is an enormous network of special cells called *neurons*. They are the fundamental processing unit of the organ that makes life as we know it possible. Like most cells in the human body they have a central nucleus, but after this common feature the neuron distinguishes itself. A thin output fiber (*the axon*) extends from the soma cell body to make contact and communicate with other axons. Microscopic arms (*dendrites*) reach behind the cell body to provide a landing area for other neurons to make contact. At these connection sites (*synapses*), terminal branches release neurotransmitters into the receiving neuron. Certain chemicals can raise the cell's default electrical charge or "excite" it while others lower or "inhibit" it. Depending on a neuron's polarity (*activation level*) it will be either more or less likely to shoot its own signal up the axon to eagerly waiting dendrite brethren. The entire process takes place in less than a fraction of a second and repeats itself millions of times each day. In this way the brain monitors, communicates, and commands the human body.

The probability of a neuronal firing (deemed action potentials) is based on the total input received at a given moment in conjunction with the neuron's default activation level. That total input is a function of the number of connections, the *weight* or size of the connections, their polarity (excitory or inhibitory), and of the *strength* of the incoming signals. Weight corresponds to the surface area in contact between the two cells and the strength of the signal can range anywhere from 0 to 200 hertz.

Individual neurons are functionally insignificant in regards to the accomplishments of the larger neural network. It is their unique organization that yields such a powerful tool. Neurons frequently arrange into a population which send their axons to the site of a second neuronal population, where the arriving axons divide into their terminal branches to make synaptic connections with multiple cells in the target population. Axons from cells in the second population of neurons project to a third population of cells, and so on (ANP, 161).

One way of studying the brain is to observe the brain itself but this approach is limited. It is too large and complex to completely map out and too vital to tinker with in a living person. Another way to study the brain is to mimic it in simpler but functionally comparable artificial networks. Below is a neuron-like processing unit called a node.

160 The Structure of Science

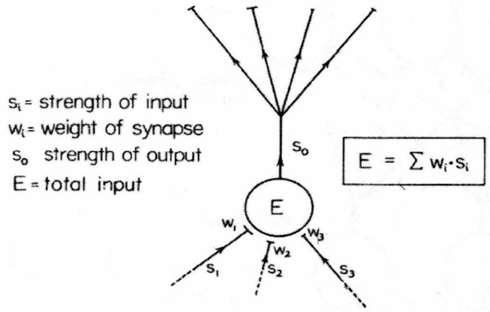


Figure 9.2
A neuronlike processing unit

Figure 9.2 (ANP, 160)

Figure 9.2 is admittedly oversimplified for the purpose of explanation. Messy curvatures have been straightened and the bushy dendrite branches have been eliminated. The structure and function of the "neuron" remains legitimate. A single node like a single neuron is explanatorily insignificant. Rather, it is their arrangement that matters. Below is a 11-node feed-forward network divided into three layers, input, hidden, and output.

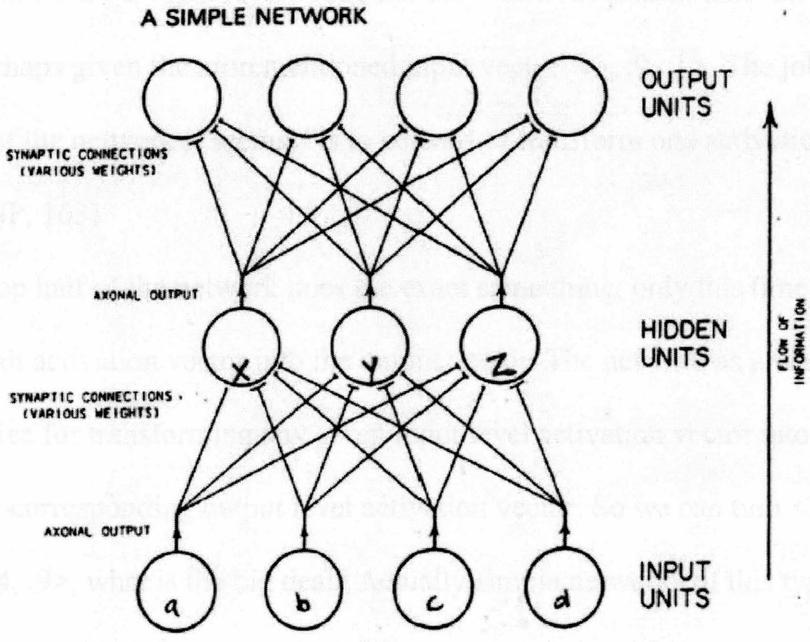


Figure 7.17 (M & C)

In this simple network the bottom layer is meant to imitate our sensory cells which receive input from the external environment. Their activation level is designed to isolate a specific aspect of the overall input stimulus striking the bottom layer. The assembled set of simultaneous activation levels at nodes a, b, c, and d is the network's *representation* of the input stimulus. Since the configuration of stimulation levels is just an ordered set of numbers, we refer to it as the input vector. To get a specific example we plug in magnitudes to the general input vector <a, b, c, d>, perhaps arriving at something like <.1, .8, .4, .5> (where the possible magnitudes range from 0-1).

The input activation levels are propagated upwards towards the middle layer of the network where each input unit has made a synaptic connection with each "hidden unit." The synaptic connection weights are varied as illustrated by the differing terminal branch ending lengths. Each hidden unit's activation level then is the weighted sum of influences sent from the input layer. Together they form the hidden unit activation vector <x, y, z>, perhaps given the aforementioned input vector <.3, .9, .1>. The job of the bottom half of the network it seems, "is to convert or transform one activation vector into another" (ANP, 163).

The top half of the network does the exact same thing, only this time transforming the hidden unit activation vector into the output vector. The network as a whole then is merely a device for transforming any given input level activation vector into a meaningfully corresponding output level activation vector. So we can turn <.1, .8, .4, .5> into <.2, .1, .4, .9>, what is the big deal? Actually simple networks of this type are capable of accomplishing some impressive feats.

Assume we are submarine engineers given the task of designing a sonar system that can distinguish between the sonar echoes of explosive mines and similarly shaped harmless rocks. Both sound indistinguishable to the human ear and both display a wide variation in sonic character. You can guess what sort of device can solve the problem.

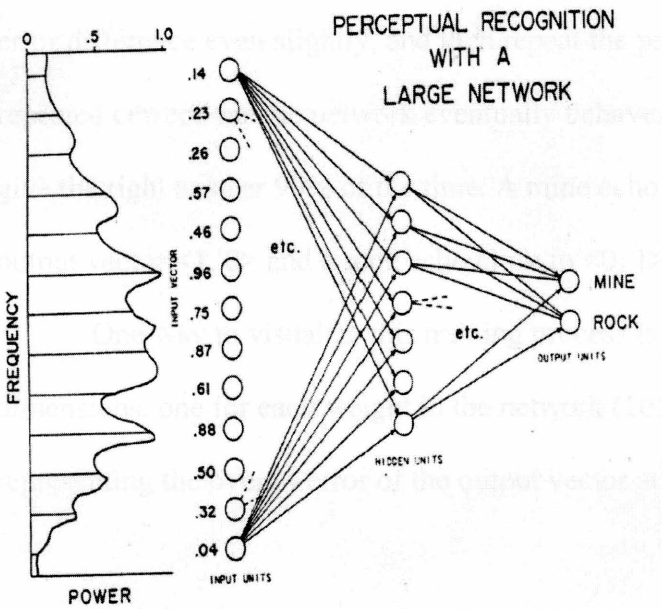


Figure 7.19 (M&C)

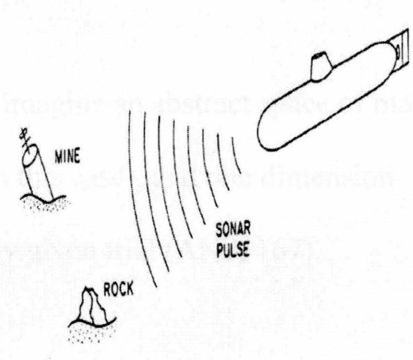


Figure 7.18 (M&C)

> This one has thirteen units at the input layer, since we need to code a fairly complex stimulus. A given sonar echo is run through a frequency analyzer and is sampled for its relative energy levels at thirteen frequencies. These thirteen values expressed as fractions of 1, are then entered as activation levels in the respective units of the input layer. From here they are propagated through the network, being transformed as they go, as explained earlier. The result is a pair of activation levels in the two units at the output later. We need only two units here, for we want the network eventually to produce an output activation vector at or near <1, 0> when a mine echo is entered as input, and an output activation vector near <0, 1> when a rock echo is entered as input. In a word, we want it to distinguish mines from rocks (ANP, 164).

The connection weights that determine the transformational activity are initially set at random values. Barring a miraculous first guess the network fails at its directed purpose, but we can teach it. Given a large recorded set of various mine and rock echoes, we can feed the network one by one with the advantage of knowing what the input is, and therefore what we would like the output vector to be. We then observe how far off the output vector is from what we would like it to ideally be, and record that difference. Either manually or via a computer training program, the synaptic weights are tinkered with slightly to identify which ones were most at fault. The hope is that we can reduce the error difference even slightly, and then repeat the process. Under the pressure of such repeated corrections the network eventually behaves how we would like it to. It begins to give the right answer 90% of the time. A mine echo produces something close to the output vector <1, 0> and a rock echo close to <0, 1>.

One way to visualize this training process is to imagine an abstract space of many dimensions, one for each weight in the network (105 in this case), and one dimension representing the overall error of the output vector at any given trial (ANP, 167).

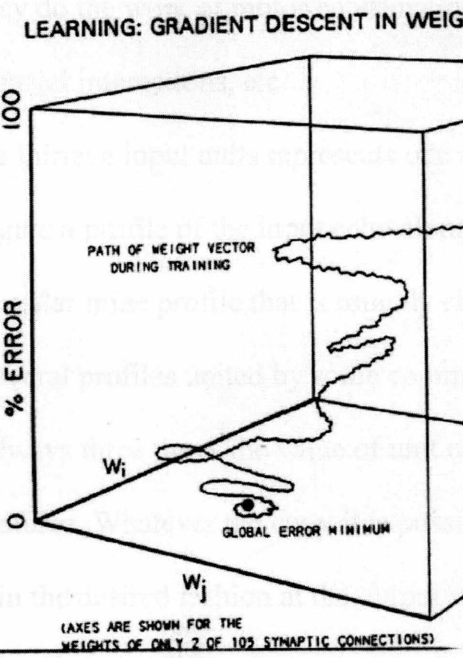


Figure 7.20 (M&C)

The above three dimensional cube is actually a 106 dimension hypercube and each point on the vast weight space represents a unique overall configuration of weights in the network. If our training is successful then the weight space point should travel down the error gradient towards a global error minimum. It is not necessary that the network will achieve its goal, that is, distinguish between mine and rock echoes. There might be no identifiable difference between the two. However in this case the network did travel down the error gradient, apparently recognizing some complex relational or structural features in the input. Once learned, that distinguishing knowledge can be applied to echoes outside of the training set. Indeed the network is only slightly less successful at identifying echoes it has never encountered before. The "knowledge" the network has acquired, "concerning the distinctive character of mine echoes, consists of nothing more than a carefully orchestrated set of connection weights" (ANP, 167). It is a fairly new and difficult concept to imagine that a set of connection weights embody meaningful knowledge, but they do the work of motor coordination, perceptual recognition, moral knowledge about social interactions, etc.

Each of the thirteen input units represents one aspect of the incoming stimulus and together they give a profile of the input echo along thirteen different dimensions. There may be a singular mine profile that is roughly characteristic of all mine echoes, or maybe there are several profiles united by some common feature (e.g., the activation level of unit 6 is always three times the value of unit number 12); or perhaps a disjunctive set of relational features. Whatever the case, it is possible to adjust the weights so that they will respond in the desired fashion at the output vector to all and only the relevant profiles (ANP, 168).

The units at the hidden layer are vital to the network's distinguishing ability. Consider a new abstract space where the seven axes represent the possible activation levels of the seven hidden units.

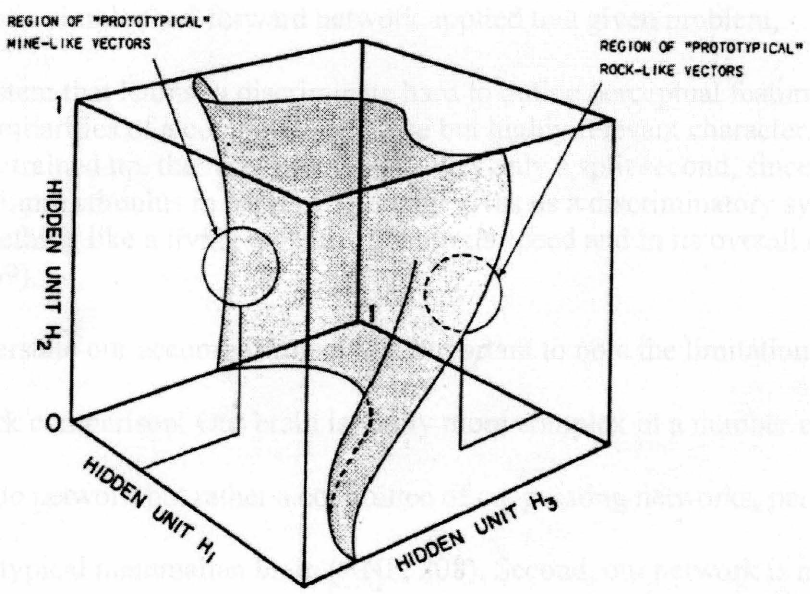


Figure 9.7 (NP)
Learned partition on hidden-unit activation-vector space. Axes are shown for only three of seven hidden-unit activation levels

What the system is searching for during the training period is a set of weights that partitions the hidden-unit activation-vector space so that mine inputs fall into a particular subvolume and rock inputs the other. The job of the top half of the network then is to distinguish between the two subvolumes which the hidden-unit activation-vector space has been divided (ANP, 168).

Looking at the above figure it appears that vectors near the center (or along some high dimensional path) of the mine vector subvolume represent prototypical mine echoes, which will produce an output vector near the desired <1, 0>. Vectors closer to the partitioning surface of the hidden-unit activation-vector space represent problematic or vague echoes, producing an output vector of something like <.4, .6>. The network's discriminative responses are not a simple yes and no, mine and rock. Rather they are graded, "the system is sensitive to *similarities* along all of the relevant dimensions, and especially to rough conjunctions of these subordinate similarities" (ANP, 168). Taking the structure of a simple feed-forward network applied to a given problem,

> We have a system that learns to discriminate hard to define perceptual features, and to be sensitive to similarities of a comparably diffuse but highly relevant character. And once the network is trained up, the recognition task takes only a split second, since the system processes the input stimulus in parallel. It finally gives us a discriminatory system that performs something like a living creature, both in its speed and in its overall character (ANP, 168-169).

Before we overstate our accomplishment it is important to note the limitations of the simple network comparison. Our brain is vastly more complex in a number of ways. First, it is not a single network but rather a committee of cooperating networks, perhaps over a thousand in a typical mammalian brain (ANP, 208). Second, our network is not the one directional feed-forward type. Input to the hidden layer comes from the sensory periphery *and* recurrent pathways – descending axonal projections from higher up populations in the network ladder.

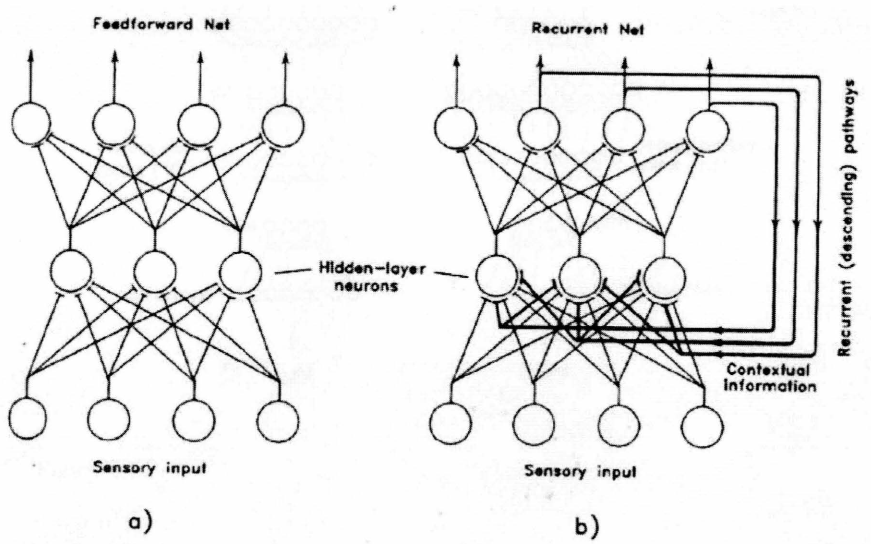


Figure 7.8
(a) A purely feedforward network. (b) A recurrent network. Note the flow of information from layer 3 back to the neurons in layer 2. This allows the network to modulate its own responses to sensory inputs, and to profit from its own immediate history.

"Recurrent networks" like our brain are able to take contextual information formed at the output layer and incorporate it into the hidden layer. As noted in the comments below the figure, "this allows the network to modulate its own responses to sensory inputs, and to profit from its own immediate history." The diagram represents recurrent pathways on top of the simple feed-forward network, but in our brain these pathways extend downwards to all layers, and horizontally from entirely different networks.

The recurrent network in figure 7.8 is oversimplified in a couple other ways as well. Networks in the brain do not consist of 3 distinct layers, input, hidden, and output. Instead, they are often over 20 layers tall (though they need not be that large), each layer with its own unique transformational purpose.

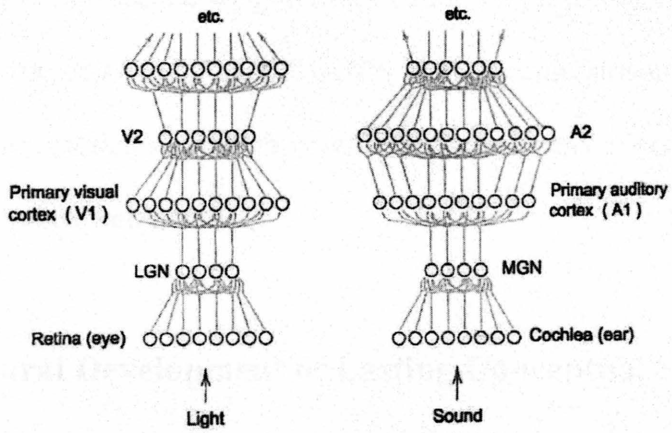


Figure 2.1

These layers do not consist of 3, 4, or even 13 neurons. Each neuronal population or rung on these 20+ step ladders are made up of millions of neurons. Finally, neurons are not uniform in structure and function. There are many types each with a role to play. Interneurons for example are the most prominent in the brain. They are the glue that holds the massive network together, receiving information from one neuron and passing it on to another. Unipolar neurons have a single process while multipolar neurons have many. Motor neurons carry information from the central nervous system to muscles and glands, usually stimulating muscle contractions (BFHB, 31).

Finally, there is no tiny supervisor in our head to coordinate the brain's various synaptic weights. The "back propagation of error" method utilized in the mine/rock network would be nearly impossible to implement on a scale as large as the human brain. On average we have 10^11 neurons, each with about 10^3 connections each, for a total of approximately 10^14 weights to adjust. If we conservatively estimate that each weight has 10 possible values, "the total number of distinct possible configurations of synaptic weights (that is, distinct possible positions in weight space) is 10 for the first weight, times 10 for the second weight, times 10 for the third weight, etc., for a total of…10^100,000,000,000,000!" (ANP, 190). By comparison, the total number of elementary particles in the observable universe is only about 10^87! So how do we sculpt this enormous weight space?

Structural Development of Lasting Conceptual Frameworks

The answer lies at least partly in a process called "Hebbian learning" named after D.O. Hebb, who first proposed the mechanism. The basic idea is that when two connected neurons fire, the strength or weight of their connection increases (BFHB, 273). Hebbian learning is a mindless, sub-conceptual process. Specifically, it is sensitive to temporal coincidences among many axonal messages arriving from an upstream population to a receiving population. If a message is repeatedly sent along a particular synapse, in other words, if both the pre-synaptic and post-synaptic neurons fire, then the connection weight between the two increases. Plainly, "neurons that fire together, wire together" (ITB, 3). Thus, the pattern of response on a receiving population gradually becomes a reliable indicator of what external feature of the world continually prompted it in the first place. Furthermore, "since the salient features in any environment are those that display a repeated pattern of development over time," the unfolding behavior of the receiving neuron over time can become a reliable indicator of a salient *causal* process of the world (ITB, 3).

An important note about Hebbian learning is that there is no supervisor. There is no "belief" being forced upon the network. There is no error message because no goal frames what is desirable and what is not. Yet Hebb rules have been shown to mold computationally impressive networks. The only requirement for this type of learning is a constant stream of sensory input. In a sense the supervisor of Hebbian learning is the external world, painting a picture of itself in our brains.

Hebbian learning is not the only type of learning we employ and there are various strategies or Hebb rules within the broader title. There are also a whole host of other strategies outside Hebbian learning that shape our cognitive landscape. Whatever the strategy, it is clear that the brain slowly shapes its cognitive landscape over time through synaptic adjustment. No one is born with effectively useful background knowledge. As already mentioned, there are approximately 10^14 weights to adjust. The entire human genome contains only about 10^9 nucleotides. To code for the exact configuration of neurons and their synaptic weights would require blatantly too much information, claiming otherwise is willful scientific ignorance. What the genome *does* do is specify endogenously the general structured principles of a type of learning network, "that is then likely to learn in certain standard directions, given the standard sorts of inputs and error messages that a typical human upbringing provides. This places the burden of steering our conceptual development where it belongs: on the external world, an information source far larger and more reliable than the genes" (ANP, 189).

Our knowledge of the world's enduring structure is embodied in the pliable (especially in the formative years) weight configuration of our neural network. This knowledge is gained at least partly by a learning process that is driven by whatever peculiar environment we reside. Furthermore there is no background rationality or propositional structure that dictates the possibility of acquiring knowledge. The only requirements are a neural network with general structural ideas and an environment to provide sensory input.

Individual Learning: Fast and Dynamical

Adjusting trillions of synaptic connections over a long period of time is not the only way we "learn." A well configured set of connections is not the only way to embody systematic "knowledge" about the world. There are not only structural changes in the brain but also "dynamical changes in the brain's typical or accustomed modes of operation" (MS 01, 14). Fast and dynamical knowledge deals with the *ephemeral* here-and-now rather than the *enduring* background conceptual framework. These immediate changes can take place in a matter of seconds rather than years, and usually involve no structural change whatsoever. However if a dynamical change proves to be beneficial it may result in future structural shift.

We can conceive of the brain's dynamical activity as a single point in the brain's all-up neuronal activation space, exploring the landscape molded by structural learning. The landscape consists of hills (the improbable spaces between likely activation regions) that the point is likely to slide off of, and valleys (the acquired prototypical activation categories) where the point tends to play. The size of the landscape is practically incomprehensible. If we assume that each neuron in the brain has ten functionally significant levels of activity, and there are 10^11 neurons in the brain, then the dynamical activity point's playground has 10^100,000,000,000 distinct possible global activation points to explore. That space cannot be even remotely explored in a single human lifetime (2 x 10^9 seconds). If our activation point covered 100 possible points every second, then we would only touch 2 x 10^11 distinct points, or about one billionth of the total space. (MS 01, 14).

The space is so vast that our point should never occupy the same activation point twice in a given lifetime. Every moment is novel. However, we normally stay in tiny sub-regions of the landscape that have their own peculiar shape. If we were to for some reason reach the top of our known valley and slide over the other side into a new sub region, we would be exploring a different peculiar structure. What pushes over these hilltops? One explanation is the reception of novel sensory input. Perhaps someone raised in a tropical climate moves to Minnesota and experiences snow and sub-freezing temperatures for the first expanding their knowledge of climatic possibilities. Another explanation is recurrent modulation. Inputs that normally traverse a specific neuronal journey are sent up a new path. The new region they enter is already molded by the slow structural learning and so contains a set of knowledge. The input is then interpreted under a new set of contextual information. For example, you are sitting in the movie theatre watching "The Sixth Sense" when (spoiler alert!) you discover that Bruce Willis' character has been dead the entire movie. Your immediate reaction is to reflect on the rest of the film in an importantly different way. It is not just the fact that you received new information but that afterwards you take known information (what it is to be dead) and apply it to a new region of already learned information (the movie prior to the twist).

In the new valley of background knowledge sensory inputs accustomed to a particular region now receive a different regime of conceptual interpretation. Perhaps the novel environment yields "an increased capacity for anticipating and manipulating one's environment, or some specific aspect of it" (MS 01, 18). Then that creature has a new insight into the world. This type of learning occurs without altering the synaptic weight configuration of your network. Though the nervous system has not changed, such a case is still a vital way of learning. Even if our weight vectors were frozen over time (they are not), one could spend a thousand lifetimes learning via redeployment of already formed conceptual resources to a new target or set of circumstances.

Collective Learning and Cultural Transmission

So far we have classified a brain's dynamical learning (trying to apply its concepts to an ever-expanding experience of the world) and the more basic structural learning (slowly shaping a useful framework of concepts in the first place). Churchland hypothesizes a third major level of learning that involves the human community rather than the individual, "the level of cultural change and collective cognitive activity" (MS 01, 21). Third level activity consists in, "the cultural assimilation of individual cognitive successes, the technological exploitation of those successes, the transmission of those acquired successes to subsequent generations, and the ever-more sophisticated *regulation* of individual cognitive activities at the first two levels of learning" (MS 01, 21-22). The various mechanisms of human culture serve to nurture, regulate, and amplify the cognitive activities of individual humans at the first two levels of learning.

Despite fathering eliminative materialism Churchland concedes that the human race's acquired public medium, language, embodies some of the acquired wisdom and conceptual understanding of the adults who use that medium. Of course the key word in that sentence is "some," it does not embody all the acquired wisdom, not even close. Enough is captured to at least provide a template of conceptual development and dynamic cognition for subsequent generations. We reap the benefits of our predecessors' cognitive achievements. Most of all, we do not have to sculpt a conceptual space from scratch. Human children learn their language from their parents and from the surrounding community of conceptually competent adults, they can shape their individual conceptual developments to conform, at least roughly, to a hierarchy of categories that has already been proven pragmatically successful by a prior generation of cognitive agents" (MS 01, 23). Language is not the only way knowledge is transmitted culturally. There are sub-linguistic practices passed on that physically alter one's cognitive development. For example if a culture is adept at a particular craft, and that craft is taught from parent to child then the child's knowledge in a specific region will be enhanced. The amplification of knowledge occurs because of communication between generations as well as communication within a culture at a given time. The specific craft is taught from parent to child, but then the child has the privilege of tapping into other children struggling with the same learning process. Where one is deficient the other may be sound and vice versa so that together they fill gaps that would not have been filled individually.

The learning process now extends beyond the lifetime of any individual. We can start from a higher point on the cognitive ladder than those that came before us because of their shared achievements. We can also reach higher up the ladder than they had the possibility of attempting. The conceptual template that the language attempts to embody can slowly evolve[1], over historical periods, to express a new more powerful view of the world (MS 01, 23). The possibility for evolution of the conceptual template means that the third level of learning, despite being orthogonal to the first two, like the first two, is plastic.

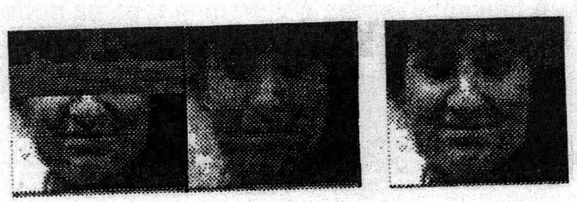
These three types of learning (slow and structural, fast and dynamical, and cultural) and their unique plasticity interact in a larger developmental story. A child raised in a specific culture will have his formative years guided by a conceptual template. The configuration of his various weightings, his enduring knowledge about the world is pushed in a particular direction. Say that child spends a year abroad in a foreign land where new received input results in a conceptual redeployment. Armed with that new knowledge upon his return he shares it with the rest of his culture. It is quite useful knowledge so the culture embraces is, and incorporates it into their conceptual template. That altered template then plays a role in the next generation's slow and structural development…etc. The plasticity at all three levels of learning work together to form our unique cognitive perspectives.

Weaknesses at all Three Levels of Learning: the Brain's Practical Imperfections

[1] As a side note it is interesting that Churchland uses the term evolve, as well as concede that the propositional sentences that make up our language do embody some knowledge. This means he does not believe all propositional categories will or should be eliminated. Although he still thinks we should move towards a more powerful framework (aka not folk psychology), he has softened his initial stance of necessary conceptual *revolution* in favor of slow *evolution.*

Hebbian learning positively defines the world's enduring structure as well as paint what the world is not. Or does it? While certain connections are reinforced, the unpopular synapses whither away. Synapses are not fixed hardware, "individual neurons compete for a given target…those synapses that are activated survive and those that are not used are eliminated; in other words, 'use it or lose it'" (Smythies, 575-576). This game of "only the strong survive" makes a lot of practical sense. Our brain power is not infinite. We have a limited amount of resources to take on a set of wildly complex tasks. It is inefficient to maintain synapses that are not used, if anything their continued presence may disrupt the network's clarity. However there are three possible explanations for unused neuronal connections. One is that the neuron was geared towards an aspect of the environment that did not exist. The second is that the organism never came into contact with the aspect of the environment the neuron was geared towards, but that environmental feature did exist. A third explanation is that the neuron did encounter the environment feature it was geared towards but only rarely. In addition when it was activated it did not provide any functional advantage to the creature. Perhaps the environmental feature was barely encountered or it was non-threatening even when encountered. In either case, the structure of our neural network similarly reacts. Thus a map is drawn of both the what and how of the enduring world, the structure both categorical and causal. However this map is only our best approximation of the world, it is not the objective world itself. Skepticism of perception advocated by Plato, Kant, and many other powerful philosophers seems appropriate here. All sensory input is inescapably interpreted through our overarching theory, but it is as they say just a theory, and a potentially misleading one.

The way we perceive the world is subjective on two levels. First, our background framework is not an objective reflection of the world itself, it is our best guess. Second, sometimes we do not have all the sensory input necessary to paint the entire picture, it is incomplete. Both our *ephemeral* and *enduring* knowledge can be deficient. Take for example a network trained in face recognition. Once mature, the network has learned a set of preferred stimuli. This background knowledge is not a memorization of the training set faces, or even a representation of something in the objective world. What is important about the preferred stimuli is that it provides, “the most effective armory for collectively analyzing any face, entered at the input layer, for subsequent placement in a well-sculpted map (the second rung activation space) of the most important ways in which all human faces variously resemble and differ from one another” (MS 02, 26). The mature network has expectations about perceptual experience and resources for placing any current input face within those background framework expectations. Given a partial face input (blindfolded in this instance), the network will repair the missing portion, re-expressing a complete face at the output layer. The diagram below shows a) the input, b) the network’s recreation at the output, and c) the woman’s actual face.



a) *b)* *c)*

This act of "vector completion" is an example of the background knowledge filling in the ephemeral knowledge gap. In speaking of what the network expects to see, "we are speaking about images that the network will *actually produce* as output, even when the input data falls objectively and substantially short of specifying, on its own, the output in question" (MS 02, 27).

What we represent is not objective reality, it is the representation best suited for navigating a difficult to perceive world. Is this means for distrust? Absolutely not, our subjective representation is more useful than what we objectively can know. The sad reality is we are perceptually limited, we cannot sense everything, nor would we want to. We lack the cognitive power to experience the entire objective world *and* solve the unique problems the objective world presents us. However by lowering our expectations of "truth" a little, by focusing on what is relevant, we seem to do quite well. Indeed the output vector for the blindfolded face input is indistinguishably close to the real full face. If a network significantly less powerful than the human brain can accomplish such a feat then perhaps we can take a little comfort in our own perceptual abilities. We may follow lies, but those lies are, at worst, quite effective navigational tools, and more often than not near the objective. The alternative of wasting cognitive power on better and more encompassing sensory equipment would be a survival blunder. The idea of not utilizing vector completion since it is merely a guess (educated by the background conceptual framework) underestimates how often that strategy is used and the theory-ladeness of all our perceptual experience.

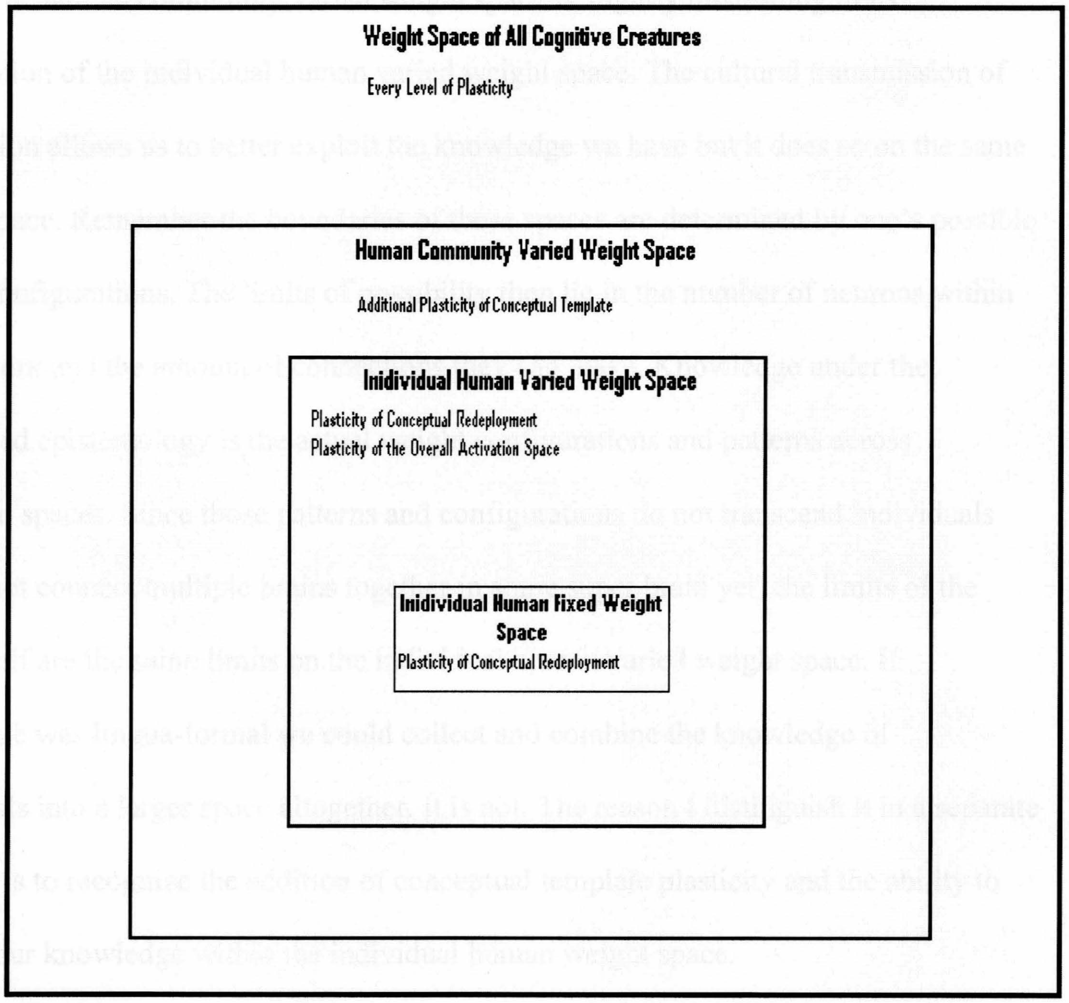
The need for vector completion or "ampliative inference" implies that we are not perfect perceptors. For example our visual system is limited, "there are a great many distinct possible reflectance profiles between which" we are unable to distinguish. We succeed because we do not aim for perfection or get overly ambitious. We choose certain relevant priorities and take care of them well. We, "forsake the impossible task of trying to compress all possible input information." Instead, we focus our limited resources on compressing the range of inputs that we most typically encounter. Fortunately, "our terrestrial environment does not display every reflectance profile, but only a comparatively small and recurring subset of them." We can deduce a simple premise from this visual compression story, limitation leads to specialization. With this premise I would like to address the concern that under the connectionist view we can not account for a common thread that unites our experience.

I have already mentioned that a conservative estimate of our weight space possibilities is 10^100,000,000,000,000 and 10^100,000,000,000 possible global activation states. Furthermore even if our weight space was frozen, "one's cognitive life might still enjoy endless novelty, simply because of the endless variety of sensory stimulation at the ladder's bottom rung" (MS 02, 3). If our brains are so malleable and configured with endless variety, one might worry that two people even raised in the same environment were too far conceptually separated to meaningfully interact. They need not worry. The answer is these numbers are mostly given for dramatic effect. Our weight space might theoretically have 10^10^14 possibilities, but in reality most of those possibilities can be ruled out.

Most weight vectors are computationally useless, under the guise of such configurations we would not enjoy the cognitive life we have now, we would not have any life, because most configurations do not work. Untrained networks take relevant information and transform it into irrelevant noise. Eliminate all of these from consideration. The synapse adjusting model of learning shrinks the space of possible global activation points, because that process makes the behavior of the higher up neurons profoundly dependent on the activities of the neurons below them (MS 01, 15). The longer we sculpt our cognitive landscape the more focused we are on a smaller area of possibilities. The space shrinks further since we all share similar cognitive and physical equipment. At the sensory periphery, our cells, whether they are auditory cells in the cochlea, or photoreceptors in the retina, are nearly identical. We all receive input in roughly the same way. As demonstrated by the reflectance profiles we are all acted upon by the same specific environment. Since we interpret the world roughly the same way, and the world acts on us roughly the same way, and, "it is clear that the world itself…is driving the learning process," it only makes sense that we all converge on a sort of isomorphic partitioning of the activation vector space even within such a vast realm of possibilities. Finally, we are all indoctrinated by a similar culture, the conceptual template is like a ball of gravity luring and perhaps restraining us into a subsection in the whole of possibilities. Together, these similarities in a typical human's experience and development place us all in a small enough section that common understanding can emerge.

Instead of focusing in on the picture, we can gain a valuable perspective by taking many steps back. Our visual system is admittedly imperfect. Although certain reflectance profiles are out of reach, we get by pretty well. Given our limitations we may have even reached a visual pinnacle, "the fact that both artificial networks and all normal humans settle into this same coding strategy suggests that it is at least a local optimum, and perhaps even a global optimum, in the space of possible solutions to the coding problems we all face." Of course the space of possible solutions to the coding problems we all face is a limited space. In the last paragraph we discovered that the conceptual realm we explore is only a tiny piece of the overall pie. The environment guides our development, but our environment is a very specific one. We live on a single planet with unique characteristics, which lies in a unique solar system, which is situated in a vast galaxy, which is only one of many galaxies in one of many galaxy clusters. Our entire known universe is arguably only a bubble in the immense cosmic foam of a multiverse. Even when we keep fixed our exact sensory and cognitive makeup there is still a space that could be explored by varying the environmental guides. On the other hand even if we fixed our environment, but had considerably more powerful sensory and cognitive equipment, we could go beyond our current boundaries of knowledge. Imagine an alien with a similar neural network as our own only with a great many more neurons, with a larger variety of functional signal strengths, and potential synaptic weights. This alien is also equipped with more sensitive cognitive equipment useful over a greater space of wavelengths and frequencies. Or perhaps this alien has a completely different network organization that we have yet to see. This superior organization is to our neural network what our neural network is to the ancient AI Turing machine. Now if we combine the broader environmental and cognitive/physical possibilities we can create a new space that subsumes our varied weight space. It is the universal weight space of all possible knowledge in all possible cognitive creatures. We might say in this realm a universal instead of a global optimum exists. We are no longer dealing with error gradients in a subsystem, we are making fundamental claims about the whole of epistemology.

Our fixed weight space cannot be fully investigated even in the combined lifetimes of a culture over history. Our varied weight space is magnitudes larger than the fixed weight space. Likewise, the universal weight space is magnitudes larger than the varied weight space.



The fixed weight space of a single individual, while the smallest section of the above diagram, is still a vast cognitive canvass. It could not be fully explored in the span of multiple lifetimes even if one desired. While the synaptic weight configurations are fixed there remains the plasticity of conceptual redeployment.

However there is no such thing as a fixed weight space in humans, the very structure of the weight space is plastic. The individual varied weight space takes into account the great many possibilities afforded by a pliable network of synaptic connections.

The human community varied weight space is slightly misleading in its subsumption of the individual human varied weight space. The cultural transmission of information allows us to better exploit the knowledge we have but it does so on the same weight space. Remember the boundaries of these spaces are determined by one's possible weight configurations. The limits of possibility then lie in the number of neurons within our network and the amount of connections they can make. Knowledge under the naturalized epistemology is the actual weight configurations and patterns across activation spaces. Since those patterns and configurations do not transcend individuals (we cannot connect multiple brains together in some super brain yet) the limits of the space itself are the same limits on the individual human varied weight space. If knowledge was lingua-formal we could collect and combine the knowledge of individuals into a larger space altogether. It is not. The reason I distinguish it in a separate category is to recognize the addition of conceptual template plasticity and the ability to amplify our knowledge within the individual human weight space.

The weight space of all cognitive creatures boasts every type of plasticity illustrated in the other categories as well as a much greater space of potential activation vector space patterns (or weight space vectors). The interplay of all three learning levels and their various plasticities takes place on an impossibly huge realm. We cannot even attempt a relational comparison to the other weight spaces because we are not sure just what type of other creatures are out there. Perhaps they have larger, more advanced neural networks of the same kind or entirely different cognitive organizations altogether. Perhaps there is no other cognitive life in the universe (or multiverse) and the weight space of all cognitive creatures is actually just the individual human varied weight space. I am doubtful but the important point is the size and structure of this space is an empirical question.

So far what we have discovered about knowledge in our small slice of the varied weight space is that it does not have a unique profile. There is no ultimate knowledge but rather a plurality. The gradient descent in weight space diagram describing the training process of the mine/rock network displayed a single path.

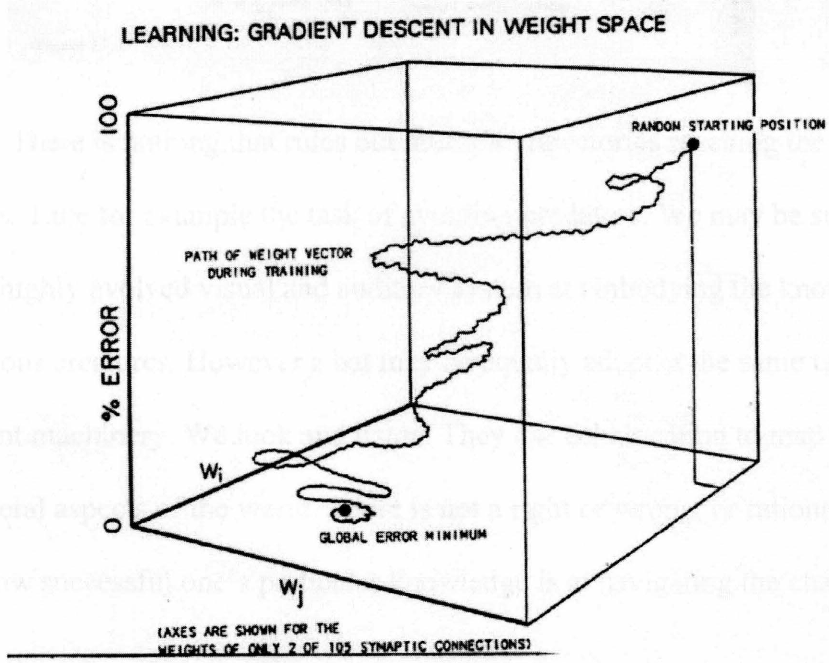


Figure 7.20 (M&C)

However that path is one of many potential ones depending on the random starting position of the weights and the nature of the input training set. Multiple global error minimum paths exist in any given task as well as multiple points on the bottom of descending error gradient. On the following network weight space only one learning trajectory reached the global error minimum but we can at least see the many potential paths one can take.

294 III. The Philosophy of Science

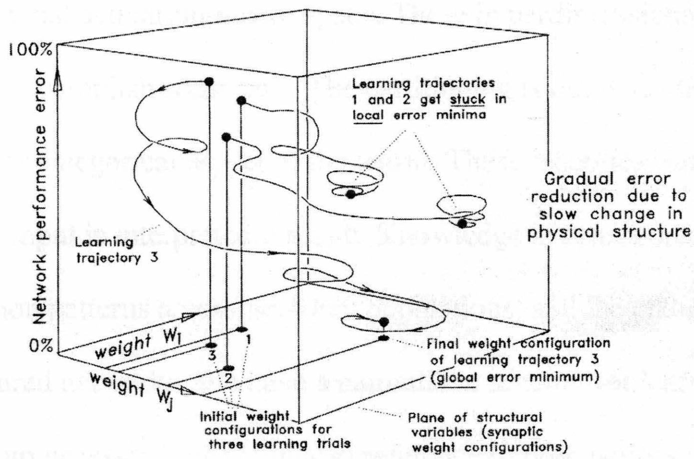


Figure 17.1

There is nothing that rules out multiple trajectories reaching the same ultimate success. Take for example the task of avoiding predators. We may be successful because of our highly evolved visual and auditory system at embodying the knowledge to avoid dangerous creatures. However a bat may be equally adept at the same task with vastly different machinery. We look and listen. They use echolocation to map the enduring and ephemeral aspects of the world. There is not a right or wrong, or rational and irrational, only how successful one's particular knowledge is at navigating the chaotic world.

To summarize the important points from the grossly oversimplified discussion of neuroscience: Our brain is a massive parallel distributive processing network. The fundamental processing unit of the brain is the neuron, which receives data and passes it on to other neurons. Groups of neurons form into networks, which consist of various layers or neuronal populations. We call the configuration of stimulation levels at each layer a vector. The job of each layer is to transform one activation vector into another in some useful way. Activation vectors can be represented as an ordered *n*-tuple of *n* quantities (e.g., <.2, .4, .1, .9> is an ordered quadruple) or as a point in an abstract *n*-dimensional activation-vector space. These hyperdimensional spaces are partitioned in a way to accomplish some task. The partitioning is our expectation or theory of some causal or categorical aspect of the world. These "theories" are the background structure that all input is interpreted through. Knowledge is then represented in the ephemeral activation patterns across neuronal populations, and the enduring structure of our configured networks. We have a naturalistic account for learning and experience that makes no necessary (or voluntary) reference to propositions. We can use what neuroscience has taught us to re-conceive many important traditional philosophical problems.

Central Themes of the A Priori

A priori knowledge is traditionally defined as knowledge that does not depend on evidence from sensory experience (Moser, 1). Other common phrases include, not empirically grounded, and necessarily true. Of course these alternative wordings all collapse back into the original. If a proposition is not empirically grounded, than it must be grounded in some non-sensory realm. If a proposition is necessarily true, then evidence from the sensory world cannot refute it. The motivation for a priori knowledge is apparent: we cannot find any true lasting knowledge from the senses, and true lasting knowledge is a goal worthy of exploration. Also, how could we acquire knowledge without some pre-existing knowledge to structure what we acquire? Therefore we should explore the knowledge that makes all other knowledge even possible.

There are many formulations of a priori knowledge, each worthy of its own paper. Leibniz said a true proposition is true in "all possible worlds." *Psychologism* advocated by Husserl claims that, "a true proposition is knowable a priori by humans if and only if our psychological constitution precludes our regarding that proposition as false." *Linguisticism* states that, "a true proposition is knowable a priori if and only if our denying that proposition would violate rules of coherent language-use," a view that denies the existence of synthetic a priori truths. *Pragmatism*, advanced by C.I. Lewis claims that, "a true proposition is knowable a priori by a person if and only if it describes their pragmatically guided intention to use a certain conceptual scheme of classification for the organizing of experiences." Roderick Chisholm asserted that a true proposition is knowable a priori by us, "if and only if our understanding that proposition is all the evidence we need to see that the proposition in question is true." For a final example, Wittgenstein asserted, "a proposition is knowable a priori by us if and only if our 'forms of life' (that is, human nature as determined by our biology and cultural history) preclude the intelligibility for us of the denial of that proposition" (Moser, 1-2).

I do not wish to tackle any of these influential thoughts in depth. That is the subject of another, probably multiple other papers. Rather, I will draw out what I see to be the central commonalities across many conceptions of the a priori:

1. The a priori is not dependent on evidence from sensory experience.
2. Knowledge is represented lingua-formally, in a set of propositions.

3. The space of all possible knowledge is the space of all possible propositions.
4. This space is stable[2] over time/location/culture, etc.
5. This space is traversable[3] in its entirety.

I wish to make the following counter claims:

1a. There is no experience outside of sensory experience.

2a. Knowledge is represented in the structure of our neural network and the activation patterns across neuronal populations.

3a. The space of all possible propositions is only a speck within a much greater area of all possible knowledge.

4a. The space of all possible knowledge is highly plastic

5a. This space is only traversable in theory, it is practically impossible.

The various conceptions of the a priori seem to converge on the external boundary or limits of all possible knowledge. They differ in how that space should be partitioned; what section is worthy of the title a priori, and what is not. I take some of the classical divides to be:

1. A priori vs. A posteriori
2. Introspection vs. External experience
3. Humans vs. Animals

I counter with:

1a. The a priori category is empty, therefore the partition is explanatorily useless.

2a. Introspection is subject to the same advantages and difficulties as direct sensory perception. The partition is relevant but not fundamental.

[2] Language can evolve over time but it only expresses propositions in a different way. The propositional structure is stable.

[3] By exercising of pure reason or conceptual analysis…etc.

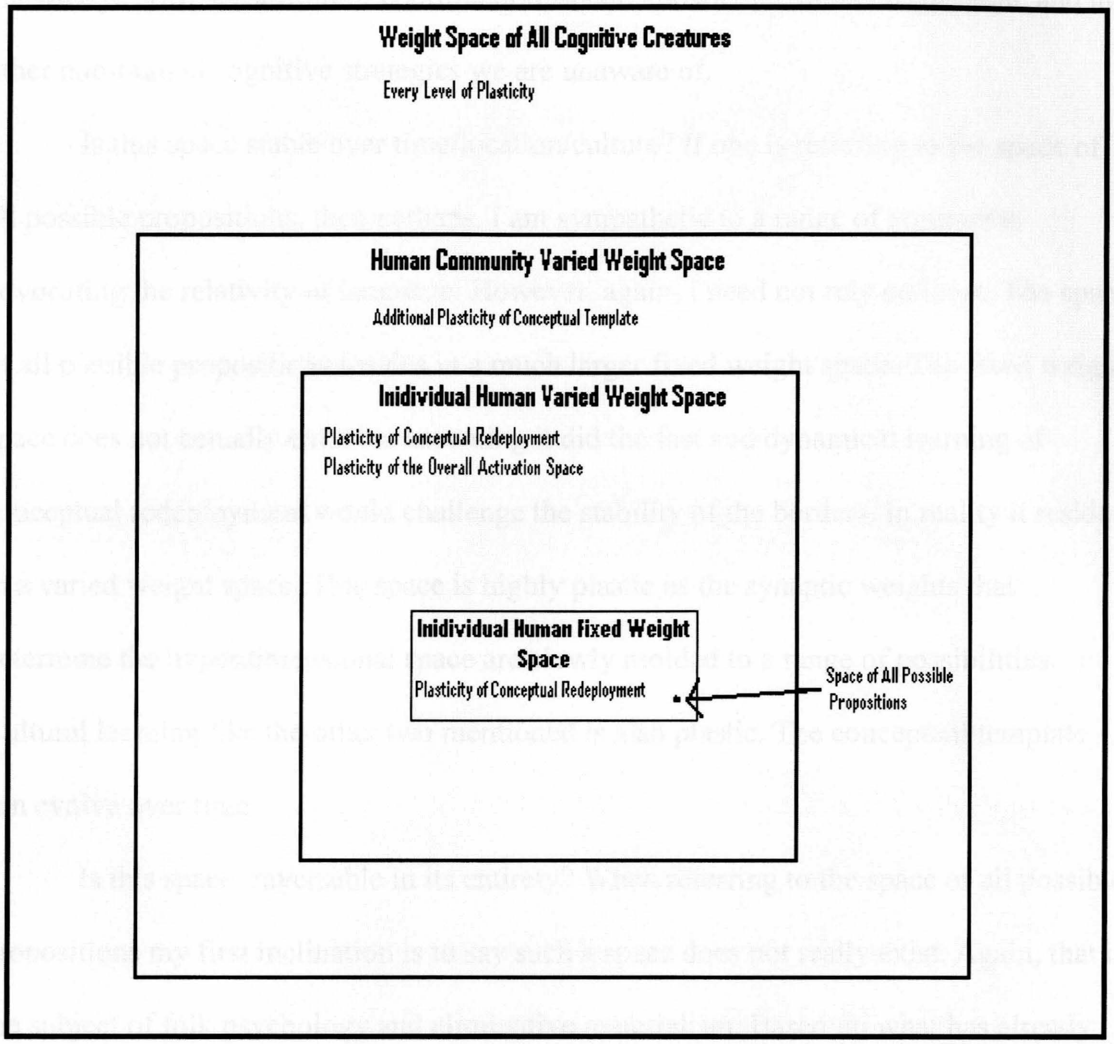
3a. Animal brains function the same way ours do, albeit with less powerful and well developed theories.

Is the a priori independent of sensory experience? Built in the question is the assumption that sensory experience is only a part of all experience. However I assert that there is no experience that sensory input does not corrupt. As the original sensory input vector travels up the processing ladder, "successively more background information gets tapped from the synaptic matrices driving each successive rung. But the 'information tapping' mechanism is the same broadly focused, wisdom-deploying mechanism at every stage. Even a purely feed-forward network, therefore, is up to its neck in knowledge-sensitive or theory-laden abductions, from its second rung representations on upwards" (MS 02, 30). Our brains are not purely feed forward networks, they are heavily recurrent, so even more background information is brought to bear on each abductive step. Potentially each step is sensitive to all of the information embodied in the entire processing hierarchy, and not just the information embodied below it in the hierarchy. And yet, the precious top of the ladder was shaped by sensory experience in the first place via synaptic modification, thus no experience is unaffected by background but that background is developed by experience.

Perhaps I am too narrowly defining experience. A common move is to claim that a priori knowledge resides in an entirely different realm. This may be the case, but then the advocates of such a move must explain how the sensible and supersensible realms interact in some causal story. We need not make this move for we already have our own coherent story that addresses for all worries within the sensible realm alone. Will we find some overarching certainty or fixity to ground all knowledge? Unfortunately not. The space of knowledge is itself plastic. Will we find the pre-existing structures that allow for the possibility of knowledge? Hebbian learning is a sub-conceptual, mindless process of synaptic weight adjustment guided by the continual inputs from one's environment. So the answer is yes, but those structures are not lingua-formal or even intelligent and are a matter of empirical inquiry.

Is knowledge lingua-formal? No. Knowledge of the general world structure is embodied in the entire activation space for the relevant population of neurons. Knowledge of the here and now is embodied in the activation patterns along a population of relevant neurons. Although in the third level of learning it is admitted that some knowledge is passed down via language from generation to generation.

A more subtle attack on the propositional conception is to agree that it does embody some knowledge but that it is an excruciatingly tiny amount. If we superimpose the space of all possible propositions within our holistic visual model of all possible knowledge, we see that it does indeed reside within the new epistemological picture.



However it is only a miniscule section of the whole. Even under some particular conception of necessity if one found necessary truths about the propositional space, those "truths" would not have any fundamental epistemological meaning. At best they would be accompanied by a whole host of other successful cognitive configurations of the world.

The diagram above answers the third common feature of a priori knowledge. The space of all possible proposition is NOT the space of all possible knowledge. It is a microscopic speck on a vast canvass of various types of knowledge: knowledge of what, knowledge how, sensorimotor coordination, social knowledge, moral knowledge, and all other non-human cognitive strategies we are unaware of.

Is this space stable over time/location/culture? If one is referring to the space of all possible propositions, then perhaps. I am sympathetic to a range of arguments advocating the relativity of language. However, again, I need not rely on them. The space of all possible propositions resides in a much larger fixed weight space. The fixed weight space does not actually exist but assuming it did the fast and dynamical learning of conceptual redeployment would challenge the stability of the borders. In reality it resides in a varied weight space. This space is highly plastic as the synaptic weights that determine the hyperdimensional space are slowly molded to a range of possibilities. Cultural learning like the other two mentioned is also plastic. The conceptual template can evolve over time.

Is this space traversable in its entirety? When referring to the space of all possible propositions my first inclination is to say such a space does not really exist. Again, that is the subject of folk psychology and eliminative materialism. Based on what has already been written here we can at least say that even if it is traversable one will have only explored an incomprehensibly small portion of all knowledge. Is the varied weight space traversable? Theoretically yes. However it would require a massive communal effort over a colossal amount of time. Furthermore many weight configurations would result in useless networks. Any individual residing in these spaces would not have the ability to discuss its position, or even contain what we consider to be a level of consciousness. The entire realm of possible knowledge involves other forms of life that we have yet to encounter and an even more mammoth space. We would need more time than the universe is old to explore it.

In terms of the partitioning, the a priori vs. a posteriori has already been dealt with. The a priori category would be empty unless one expanded the definition of experience to include happenings in some supersensible realm. Such a move is unnecessary.

The introspection vs. external experience is an interesting partition because it reflects on the heart of a stereotypical philosopher. We imagine an old man with a long beard sitting in his rocking chair by the fire simply exploring the recesses of his mind. This notion is romantic but uninformed. The slow and structural learning is a random process that takes a relatively blank canvass and using sensory input molds it into a powerful tool. Even if we were able to close our eyes and shut off every single sensory neuron what would be left is a space that was fundamentally made by sensory experience. The Platonic tradition of looking past the noise… "involves not one, but a succession of distinct steps, only tens of milliseconds apart, each one of which exploits the relevant background knowledge embodied in the peculiar cadre of synapses there at work, and each one yields a representation that is one-step less stimulus-specific, one step more allocentric, and one step more theoretically informed than the representation that preceded it in the processing hierarchy" (MS 02, 29). We have confidence higher up the ladder where the supposedly unique rationality occurs, because it is supposedly devoid of messy stimulus input. Yet, the higher up hierarchical levels are molded by sensual experience, and even after being molded, each new sensory input is incorporated into the more general contextual background. If you distrust the bottom of the ladder, you better distrust the top too!

The human vs. animal is quite distinct under a lingua-formal conception of knowledge. We can manipulate language and they can not. While they do communicate, it is not in the advanced propositional manner that we do. However the division is not as distinct under a naturalized epistemology. The knowledge of language use is an acquired skill in the same vein as a crab learning to grab food with its claws (MS 02, 9-12). Both are done by slowly sculpting a relevant cognitive space. The knowledge of each activity is embodied in the same activation vector space patterns. Animal theories are much simpler than ours, as well as less coherent, organized, and informed. However, their theories are born of the same processes, and the knowledge they contain are embodied in the same mechanisms.

The principle epistemological distinction on the naturalized view appears to be that between the ephemeral and the enduring. The former deals with the fleeting experience of here-and-now and is represented in the activation patterns along a population of neurons. The latter focuses on the lasting structure of the world and is represented in the synaptic weight configuration of the entire system. Ephemeral knowledge is inescapably funneled through the larger enduring knowledge of our weight space. At the same time ephemeral knowledge can adjust the lasting synaptic weight configurations. There is plasticity at both levels and interactions between the two. They can not be functionally separated and no cognitive process takes place without accessing these two categories of knowledge. The distinction is not propositional and is a matter of empirical inquiry.

Does this mean we should discard the entire methodology of a priori philosophy? If a priori examination requires a propositional conception of knowledge, then I would argue yes. Conceptual analysis is not a completely futile exercise, for it does reside in the holistic picture of knowledge. However, if given the choice between exploring that tiny space vs. a much larger meaningful space, I would have to choose the latter. Perhaps what we need to continue is the motivation behind a priori philosophy. That is, finding the grounds for the possibility of something, in this case knowledge. What are the grounds for the possibility of our type of cognitive experience? First a neural network that can embody knowledge about the world's enduring structure and ephemeral here-and-now situations. Next we need an environment that provides sensory input to guide the relatively blank initial cognitive canvass. From this point a complex brain can be constructed from sub-conceptual processes. Hebbian learning does not operate in some propositional overarching structure. It simply reinforces what the environment tells the brain. Finally to reach more advanced capabilities of knowledge amplification we need a community of cognitive creatures operating under a general cultural template. But of course the nature and function of these pre-knowledge requirements are empirical questions. One would have to abandon the notion of prior to experience. If anything the pursuit of the fundamental in terms of knowledge seems to be a rededication to the natural sciences, particularly neuroscience.

Works Cited

1. Churchland, Paul M. A Neurocomputational Perspective. The MIT Press. Cambridge, MA. 1992 (ANP)

2. Churchland, Paul M. MS 01 (not yet published) (MS 01)

3. Churchland, Paul M. MS 02 (not yet published) (MS 02)

4. Churchland, Paul M. Into the Brain: Where philosophy should go from here (ITB)

5. Churchland, Paul M. Matters and Consciousness The MIT Press. Cambridge, MA. 1988 (M+C)

6. Wilson, Josephine F. The Biological Foundations of Human Behavior. Wadsworth Publishing. 2002 (BFHM)

7. MOSER, PAUL K. (1998). A priori. In E. Craig (Ed.), *Routledge Encyclopedia of Philosophy*. London: Routledge. Retrieved April 14, 2007, from http://www.rep.routledge.com/article/P001 (Moser)

8. Smythies, John. The Biochemical Basis of Synaptic Plasticity and Neurocomputation: A New Theory. Biological Sciences. The Royal Society. 1997 (Smythies)

(I apologize for the rough citations my books were locked in the library)