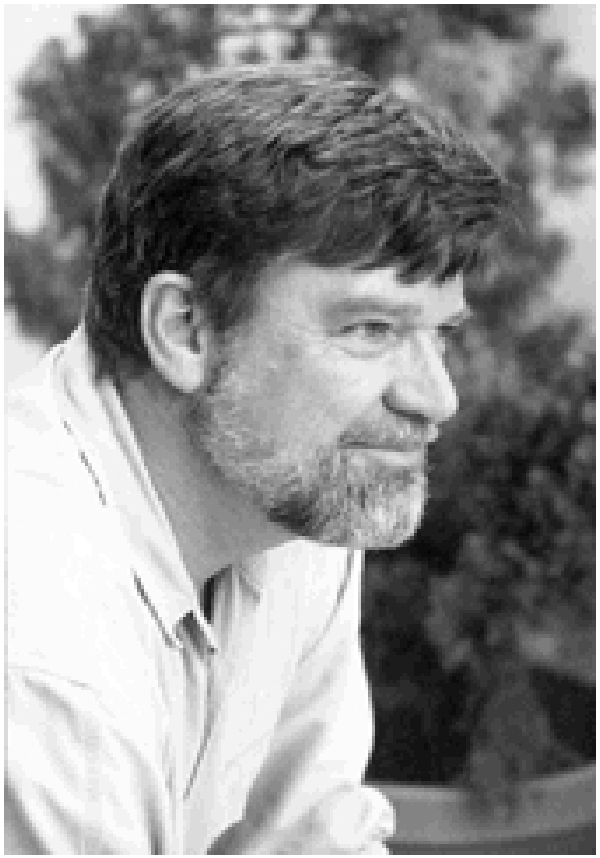
A neurocomputational perspective

Christian Wüthrich

http://philosophy.ucsd.edu/faculty/wuthrich/

15 Introduction to Philosophy: Theory of Knowledge
Spring 2010

Paul Churchland (*1942): eliminative materialism



- BA U of British Columbia, PhD Pitt (1969)
- taught at Toronto, Manitoba, UCSD
- important contributions to philosophy of mind, epistemology, perception, philosophy of cognitive science
- collaboration with Patricia S Churchland, his other “hemisphere”
- eliminative materialism:
  1. folk psychological concepts such as beliefs, feelings, desires, etc lack a coherent definition
  2. don’t expect them to be part of a strictly scientific understanding of cognitive activity, as they have no neural correlates
  3. eliminativism about propositional attitudes

What is a theory? The traditional answer

Paul Churchland, *A Neurocomputational Perspective*, MIT Press 1989, Chs. 9 and 11.

- classical account sees a theory as a set of sentences or propositions (“sentential view”)
- rationality is defined by the proper set of formal rules taken from logic
- in the sentential view, ultimate virtue of thys is truth
- Churchland: sentential epistemologies are impoverished
- Fundamental assumption of traditional view: language-like structures constitute basic, most important form of representation in cognitive creatures
- Correlative assumption: cognition consists in the manipulation of these basic structures
- cognitive neurobiology and artificial intelligence offers, according to Churchland, an alternative “framework that owes nothing to the sentential paradigm of the classical view” (154)

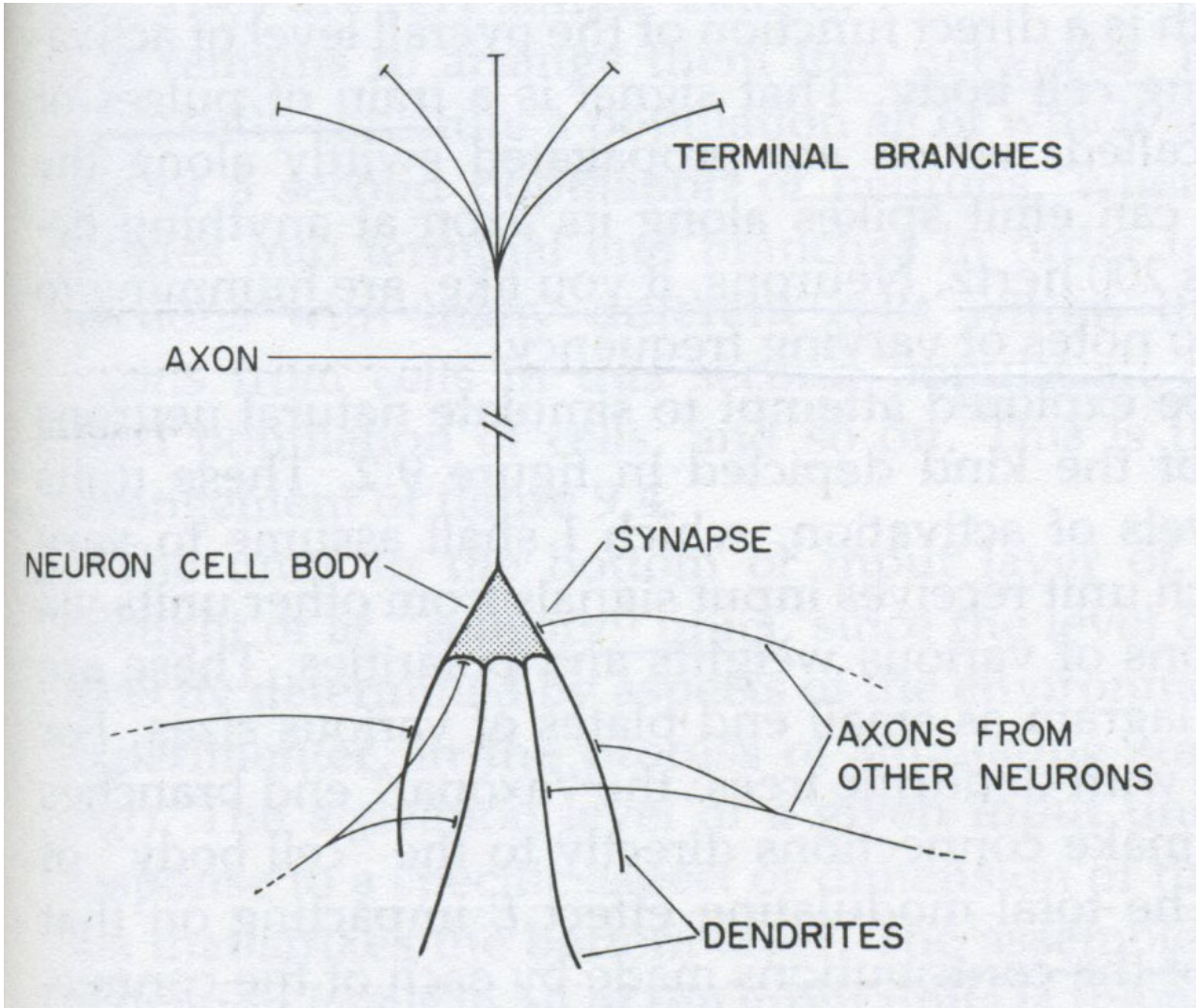
Problems of the traditional account

1. prob with infants: presupposition of propositional system and capacity to manipulate it following determinate rules fails bc this is precisely what an infant lacks prior to extensive learning
2. nonhuman animals: none of them seem to have benefit of language (although some forms of signalling), yet they clearly learn and know
3. frame problem of AI: if knowledge is storage of immense set of sentences, then retrieval and manipulation would take much longer than it does
4. knowing-how v. knowing-that: connection bw learning of facts and learning of skills cannot be made in traditional account when these two modes are inseparable in science (and more mundane contexts)

5. neurobiological embedding: constraint on any epistemological thy is that they make contact with neurophysiological accounts of how brain works is violated by trad account
6. prob of convergence to truth: while thys have become dramatically better in many respects, it is problematic to think of them as "converging" to the "Truth"

$\Rightarrow$ alternative approach: neural networks!

A schematic neuron

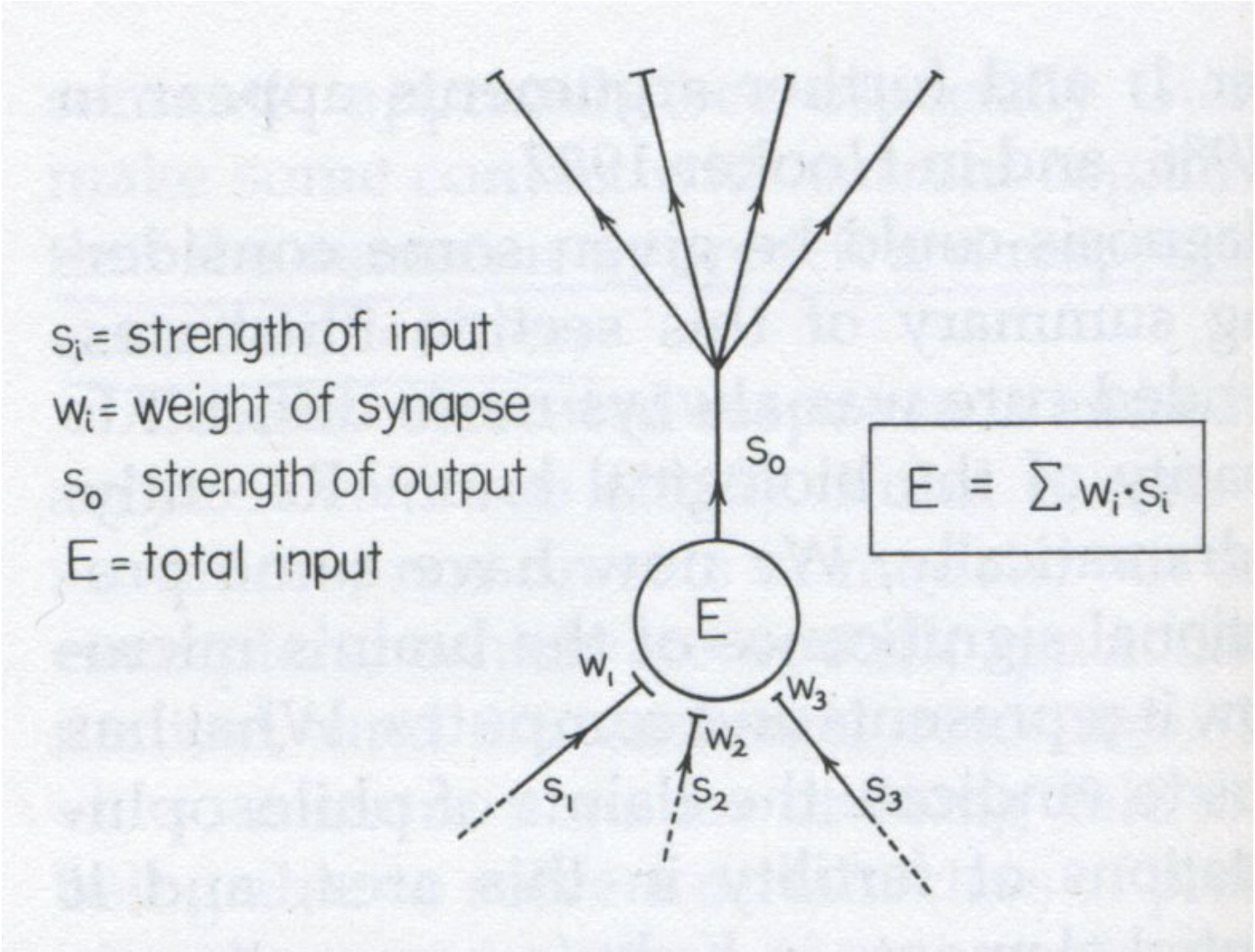


- output from each neuron: axon, makes large number of synaptic connections to other neurons (i.e. their cell bodies or their dendrites)
- input to each neuron, which is either excitatory or inhibitory
- induced level of activation is function of number of connections, from their weight, their polarity (stimulatory or inhibitory), and strength of incoming signal
- output is function of level of activation of neuron
- signals are trains of pulses (with frequency of up to 200 Hertz)

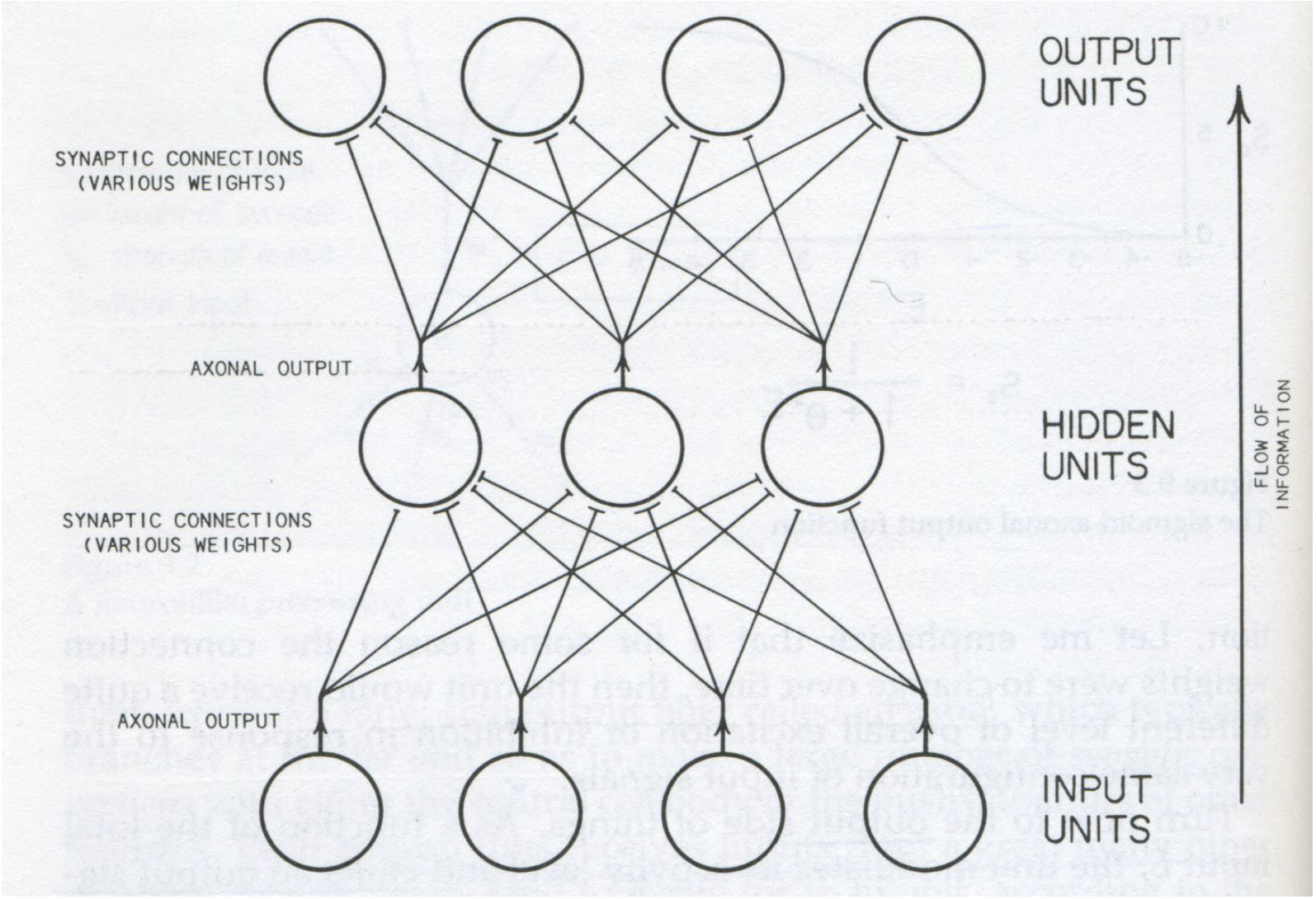
$\Rightarrow$ simulate natural neurons with artificial processing units

$\Rightarrow$ artificial neural networks

A processing unit, neuron-like node of artificial network

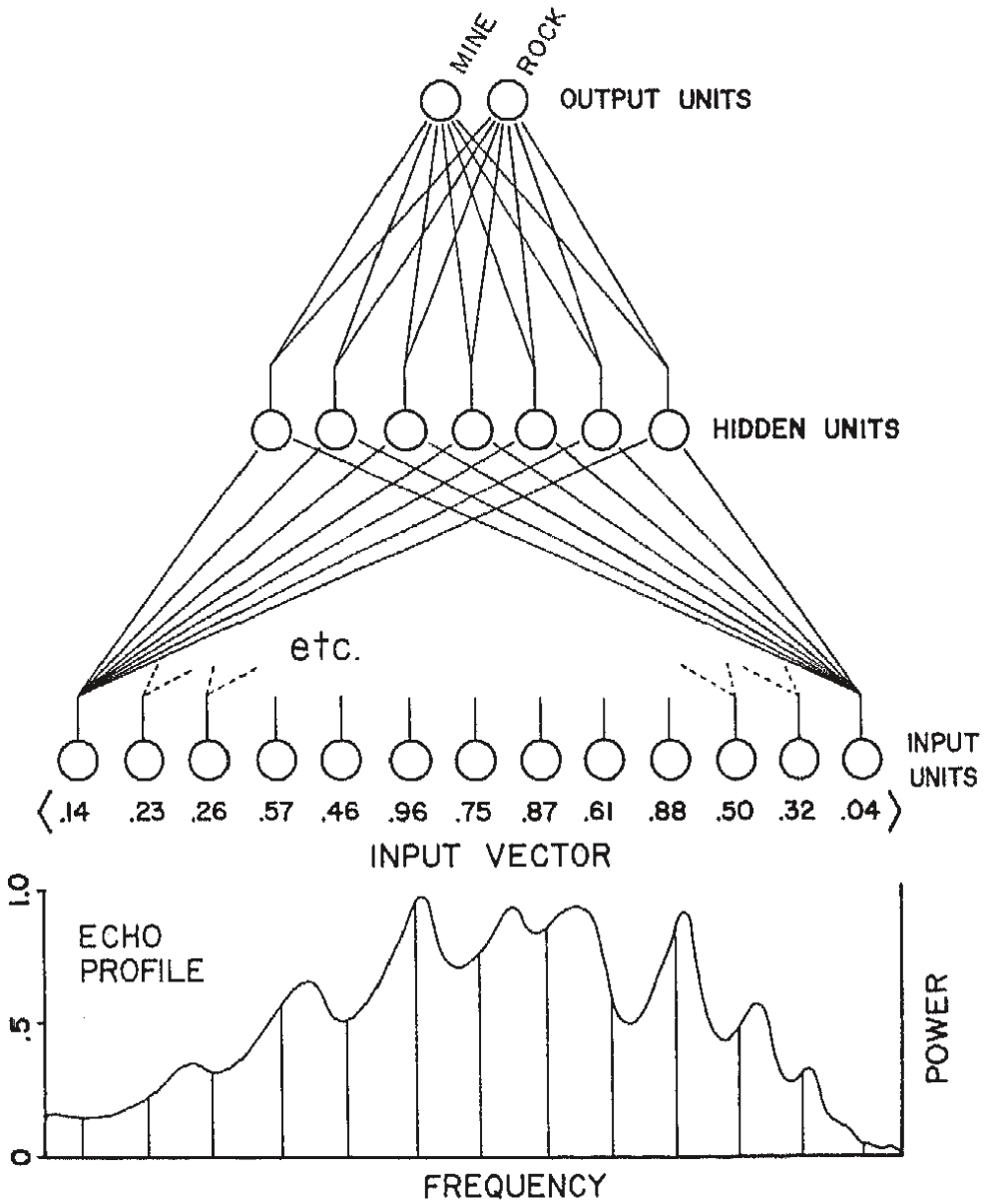


Arrange neuron-like units into networks



Neural networks as functions

- neural networks are functions which map an input vector to an output vector
- input vector: set of simultaneous activation levels in input units
- this input vector is the network's representation of the input stimulus (see figure on next page)


MINE
ROCK
OUTPUT UNITS
HIDDEN UNITS
etc.
INPUT UNITS
.14 .23 .26 .57 .46 .96 .75 .87 .61 .88 .50 .32 .04
INPUT VECTOR
1.0
.5
0
ECHO PROFILE
POWER
FREQUENCY

Neural networks as functions (continued)

- example: $\langle 0.14, 0.37, 0.59, 0.11 \rangle$
- "hidden units": middle layer(s) of network, w/ activation vector at each level of the middle layer(s)
- values of this vector uniquely determined by input vector and by the connection weights at end of terminal branches of input units
- output vector is then produced, again uniquely determined by activation vector at (highest) hidden layer, and the relevant connection weights
- output vector represents activation levels of units at output level
- in total, network is device that transforms given input-level activation vector into unique output-level activation vector

Representation in brainlike networks

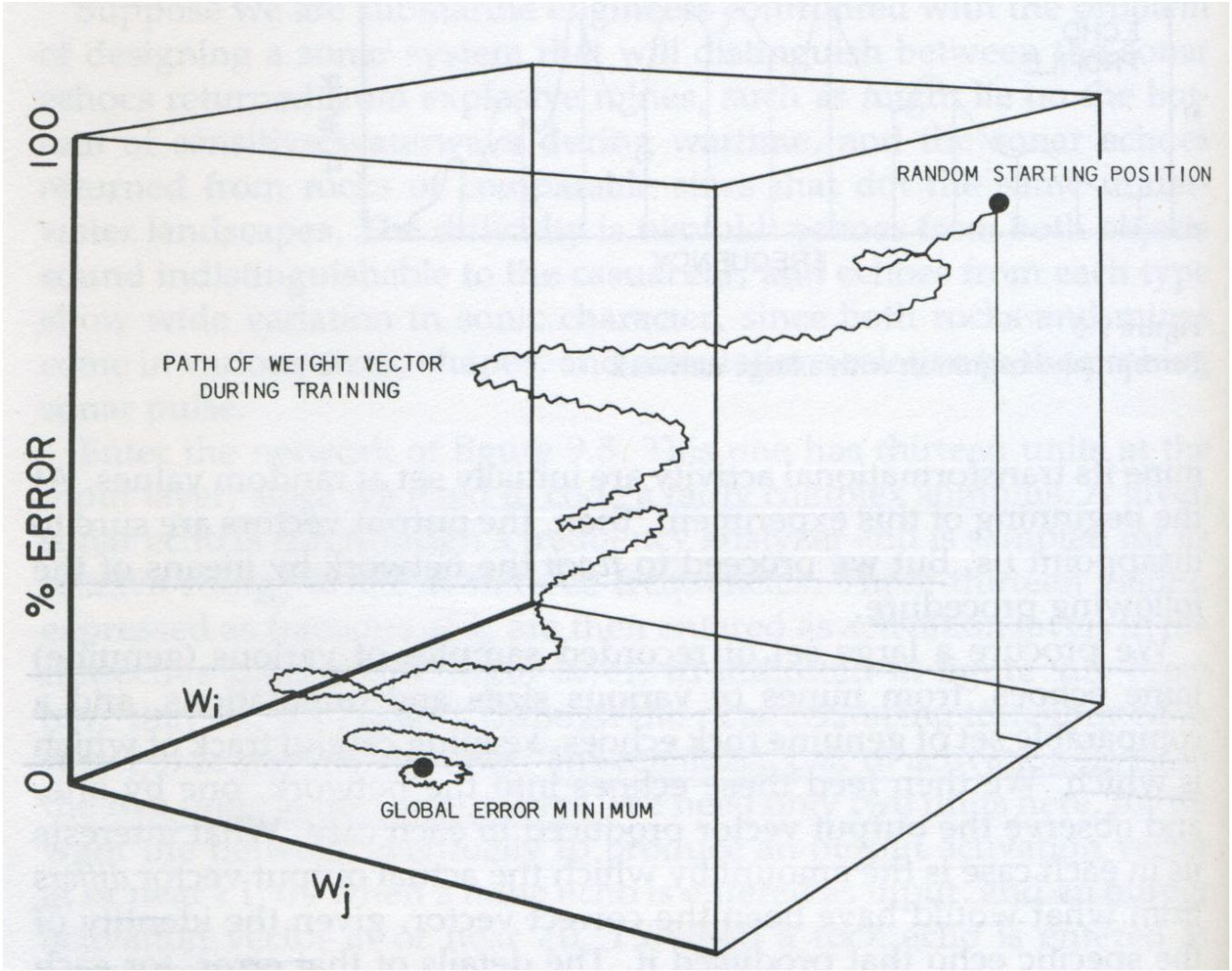
- example: vowel sound /eɪ/ as in “the rain in Spain stays mainly in the plain”
- problem: huge range of acceptable (and recognizable) pronunciations
- task of brain is to correctly ascribe same meaning to all these different sounds (or to most of them)
- same is true for colours, faces, flowers, animals, voices, smells, songs, feelings, words, meanings (including metaphorical), etc
- amazing: brains can do this!
- example: how a brain learns to tell a rock from a sea mine given a sonar echo
- difficulties: echoes from both sound indistinguishable to untrained ear; both show wide sonic variation

Of Mines and Rocks

- (look again at previous figure, three slides ago)
- task: reliably produce correct output vectors $\langle 1, 0 \rangle$ for mine and $\langle 0, 1 \rangle$ for rock
- give network a highly varied training set of examples of both
- special learning rule: computes set of small changes in values of all synaptic weights in network
- idea: to identify those weights most responsible for the error

$\Rightarrow$ "teacher" that "trains up the network"

Learning: gradient descent in weight space



So what's knowledge?

- knowledge is nothing but a "carefully orchestrated set of connection weights" (= point in individual's synaptic weight space)
- if learning is successful, synaptic weights will be such that generalization beyond training set is reliable (but not infallible)
- abstract space of *n* axes (where *n* is the number of units at the hidden level) representing possible activation levels for each of the *n* units is called hidden-level activation-vector space
- hidden-level activation-vector space is partitioned into regions of prototypical "rock"-like and prototypical "mine-like" vectors
- output level reads off from hidden level in which region of partition the hidden-level activation vector is
- if knowledge is understood in this way, it's radically different from the sentential knowledge epistemologists, philosophers of science, inductive logicians etc have traditionally ascribed to us!

Functional properties of brain-like networks

1. hidden layers in network allow for complexity, particularly powerful in recognizing regularities
2. non-linear response profile; these two properties together imply that virtually all possible non-linear trafos can be computed by network
3. Generalized Delta-rule: teaching rule for adjusting the weights, “learning by the back-propagation of error”
    - although no guarantees exist, this rule is surprisingly effective in guiding network to global error minimum (particularly if the weight space has many dimensions)

From this, Churchland concludes that

> *"it is plain that [the rock/mine network has] contrived a system of internal representations that truly corresponds to important distinctions and structures in the outside world, structures that are not explicitly represented in the corpus of [its] sensory inputs. The value of those representations is that they and only they allow the networks to "make sense" of their variegated and often noisy input corpus, in the sense that they and only they allow the network to respond to those inputs in a fashion that systematically reduces the error messages to tickle. These, I need hardly remind, are the functions typically ascribed to theories... An individual's overall theory-of-the-world... is not a large collection or a long list of stored symbolic items. Rather, it is a specific point in that individual's synaptic weight space." (177)*

How does this solve the problems listed at the outset?

- infant and nonhuman animal cognition operates in essentially the same way as human adult cognition
- massively parallel processing $\Rightarrow$ no problem with speed of relevant access (i.e. no “frame” problem)
- knowing-that and knowing-how are essentially of the same kind

How accurately do these networks depict real brains?

A few numbers...

- real brains: networks of perhaps a thousand (10^3) smaller networks
- total of 10^{11} neurons with 10^3 connection on each for a total of 10^{14} synaptic connections
- if each synapse admits of 10 distinct weights (a gross underestimate), the there are $10^{10^{11}}$ or $10^{100,000,000,000}$ distinct possible configurations of weights for each subsystem alone...
- comparison: number of elementary particles in universe (incl photons) is roughly 10^{87}

In general: overall architecture of brain is rather accurately depicted, but many details problematic...

- not all axons link to all neurons on next level
- horizontal connections within layers
- real neurons cannot change their “signs”—they’re either stimulatory or inhibitory
- serious: implementation of generalized delta rule is in computer outside network, i.e. it is far from clear that brain answers to demands of back-propagation algorithm

- delta rule presupposes a representation of what would have been correct output vector, but real creatures usually lack such perfect information
- but their brains still learn, so there must be a different learning rule
- Nota bene: innate knowledge not plausible on this view bc entire human genome contains about 10^9 nucleotides (= structural units of DNA) but would have to code 10^{14} connection weights (even though some of this may be coded recursively), and variations among individuals not accounted for by innateness

Addendum: Weight and activation-vector spaces

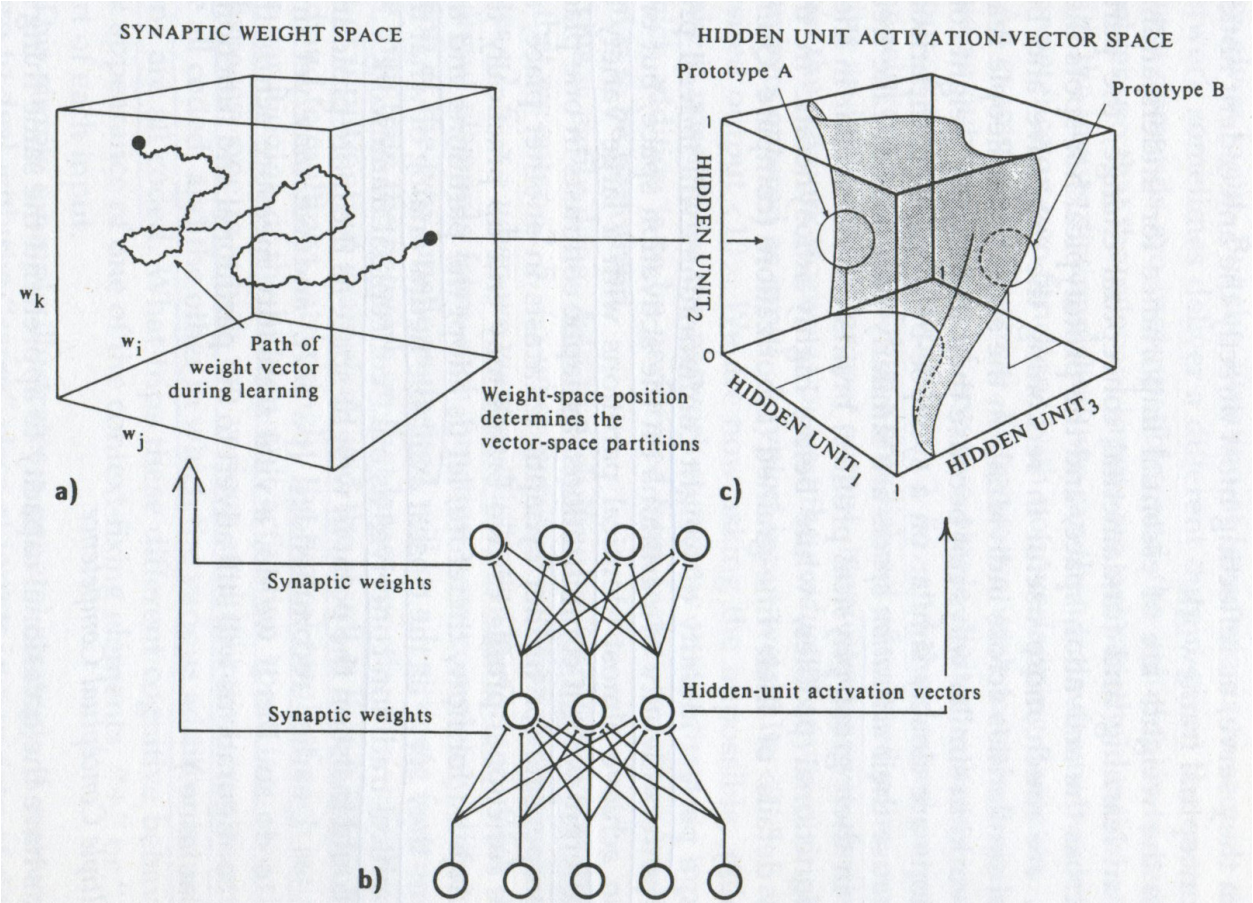


Figure 11.1
a) Weight space: the space of all possible configurations of the network's synaptic weights. Axes are shown for only 3 of 27 synaptic weights. *b*) A schematic network. *c*) Hidden-unit activation-vector space, the space of all possible activation vectors across the population of hidden units

Addendum: Mines and rocks again

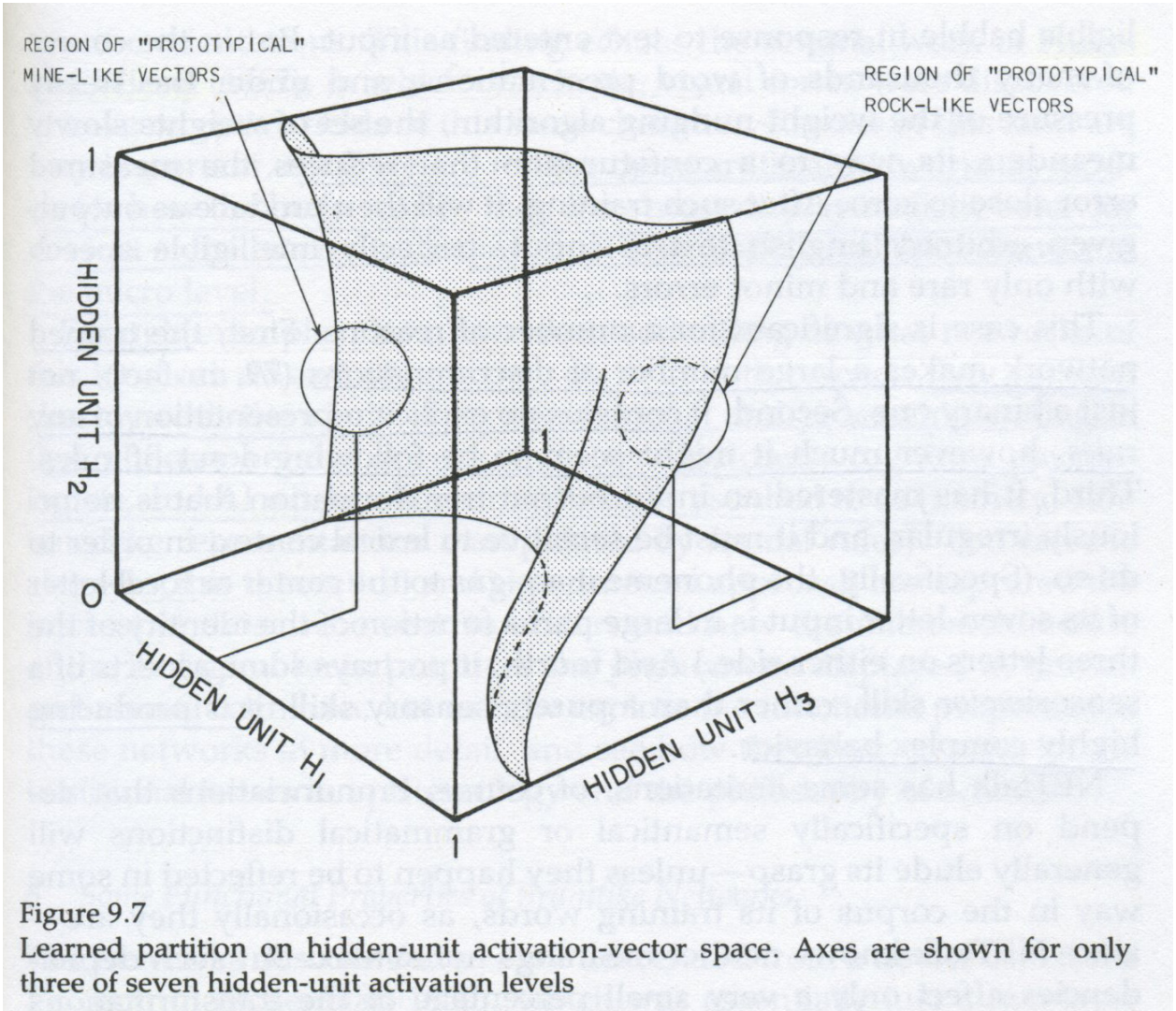
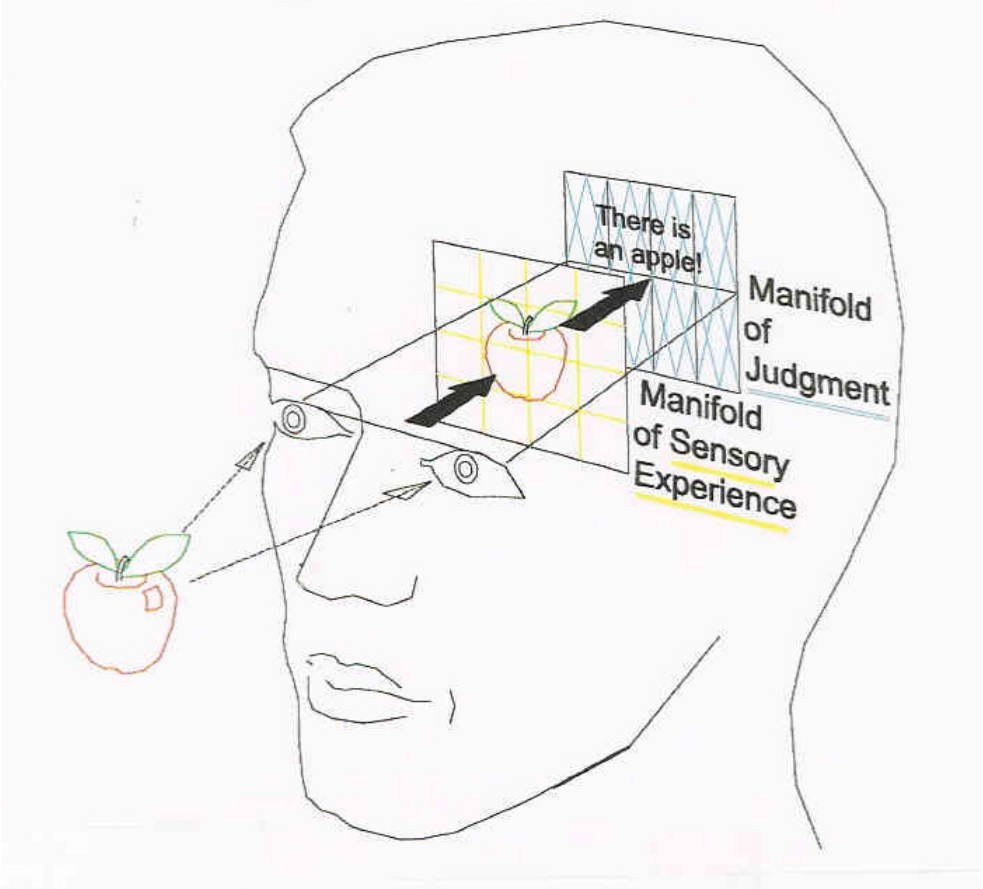


Figure 9.7
Learned partition on hidden-unit activation-vector space. Axes are shown for only three of seven hidden-unit activation levels

Networks as internal cognitive spaces

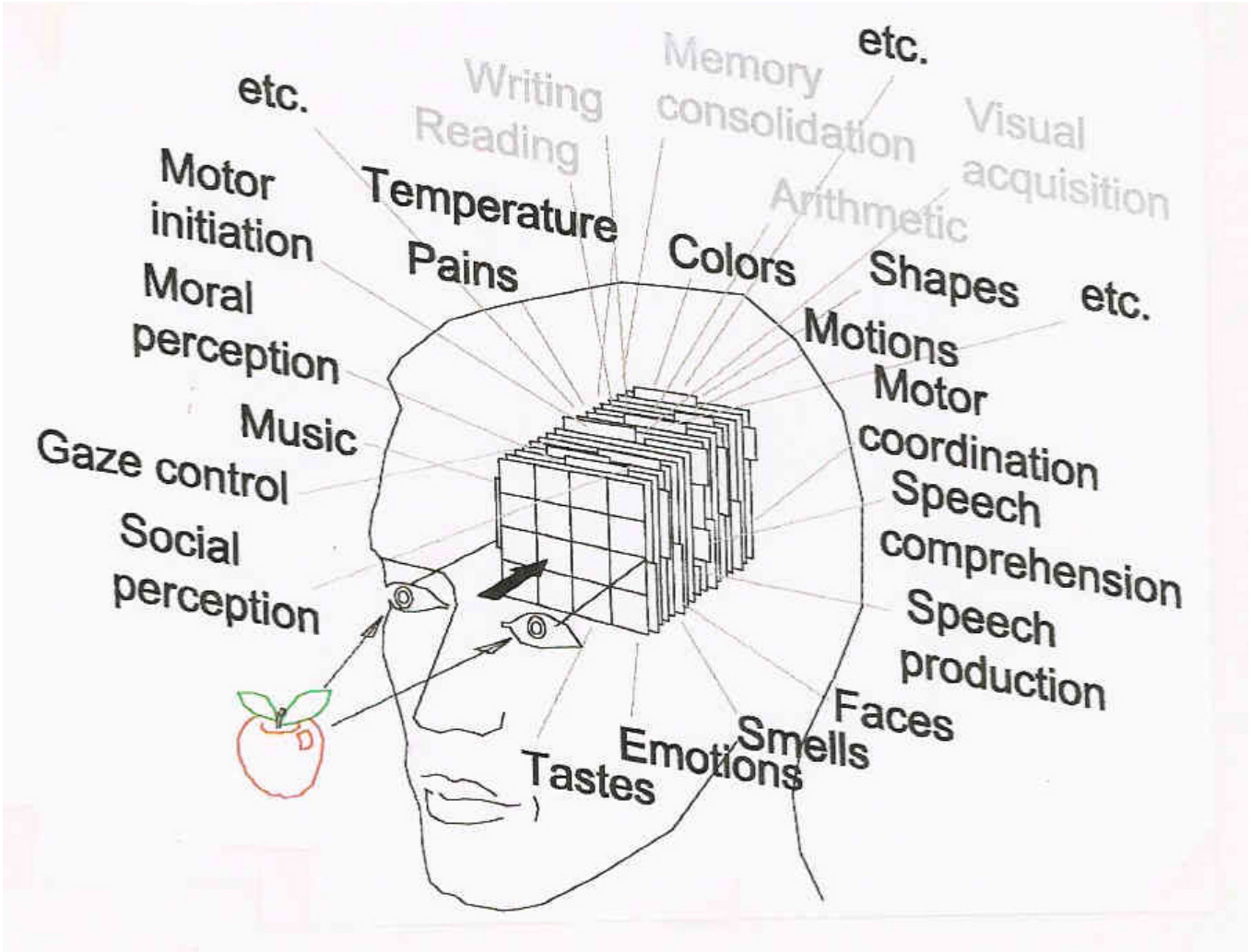
Paul Churchland (2010), *Plato's Camera*, Ch. 1.

Churchland rejects as simplistic the traditional Kantian view of cognition as divided between empirical intuition and rational judgment:

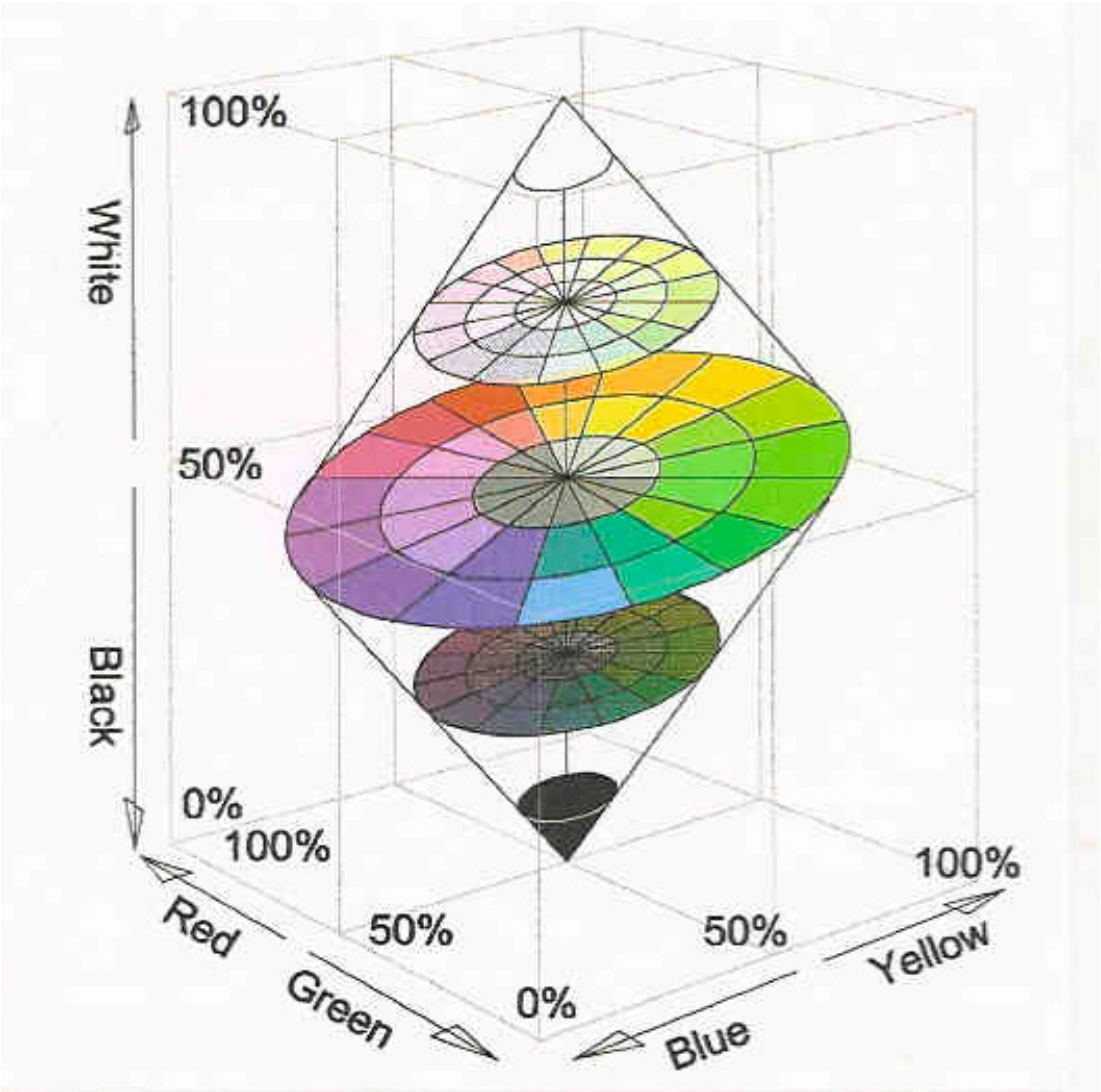


Networks as internal cognitive spaces

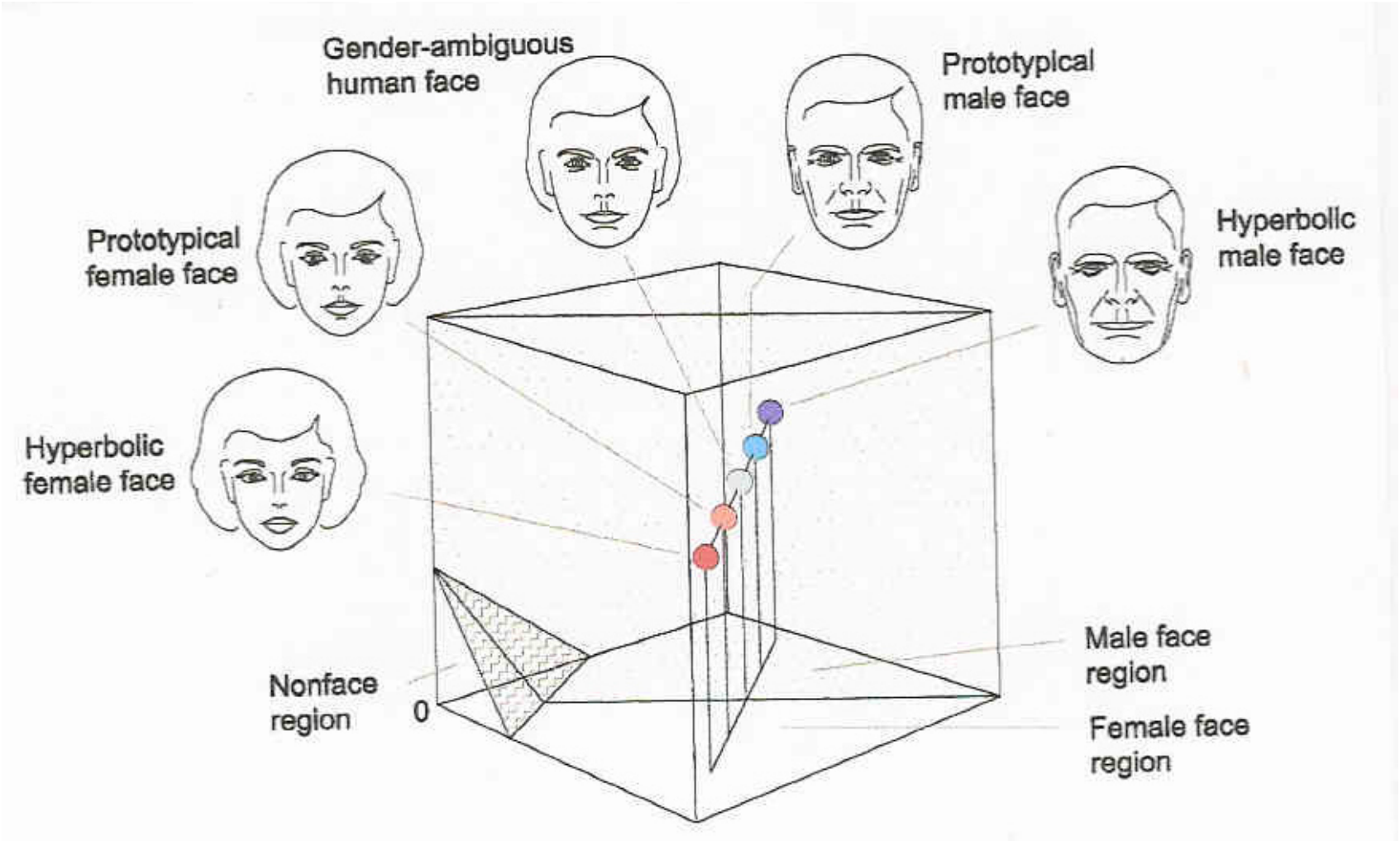
Instead, hundreds or thousands of internal “cognitive spaces”:



Example of internal cognitive space:
Space of possible colour experiences



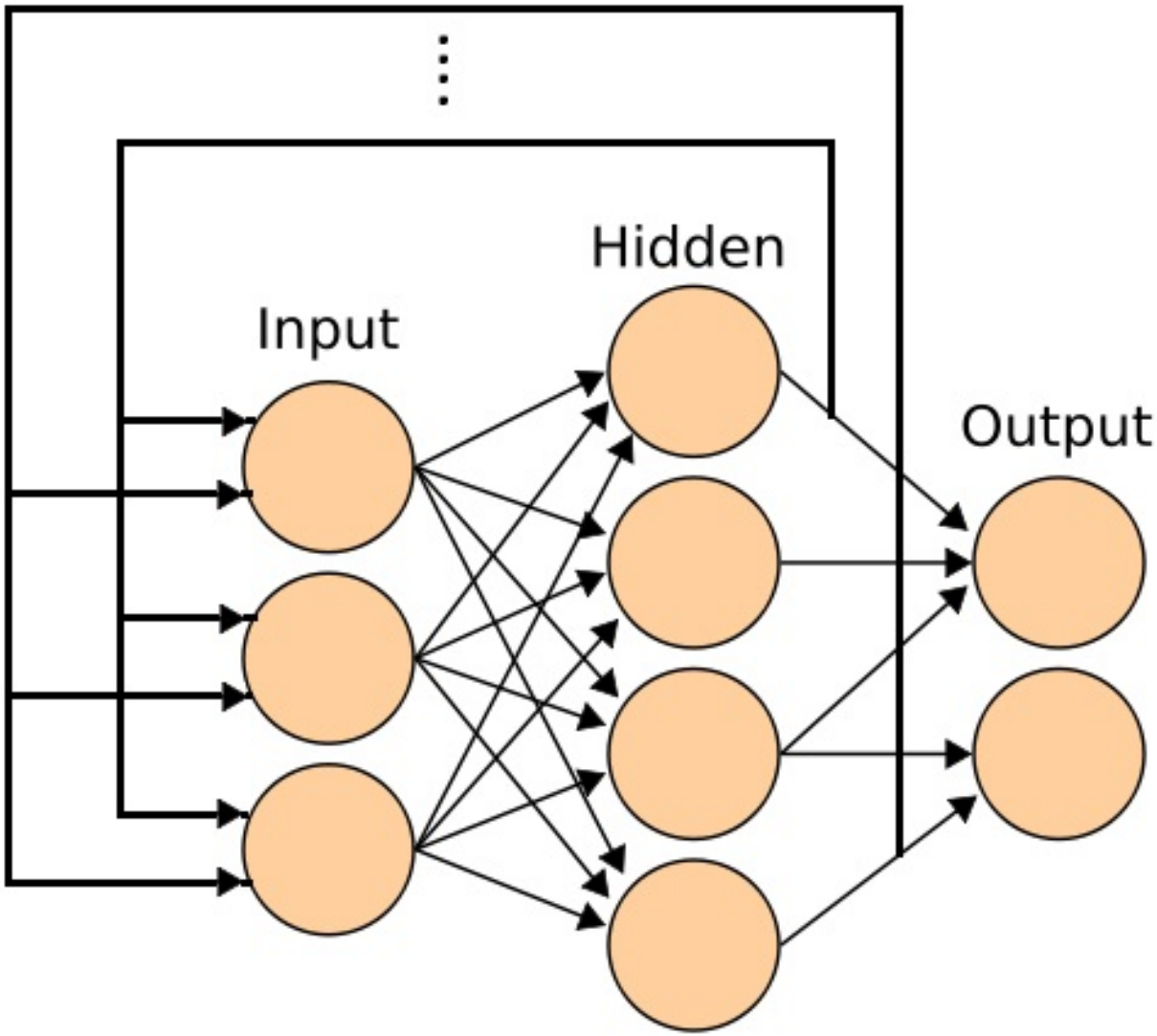
Example of internal cognitive space:
Space for representing human faces



Three levels of learning

1. Structural learning: individual, slow
   - neural networks are plastic: growth, extinction, and modification of synaptic connections, i.e. weights are adjusted over time
   - synaptic connections serve as "brain's elemental information processors, as well as its principal repository of general information about the world's abstract structure" (11)
   - More on this in a minute...
2. Dynamical learning: individual, fast
   - recurrent networks (i.e. networks with descending pathways make brain genuinely dynamical system, large range of behaviours, which are largely unpredictable even in principle
   - "ongoing modulation of brain's cognitive response to its unfolding sensory inputs" (17)
   - "conceptual revolutions" in science as cognitive ex(h)aptations, i.e. the redeployment of cogntive capacities to new tasks or in new ways

Recurrent neural network



Three levels of learning (continued)

3. Cultural and linguistic transmission: collective, very slow
   - collective medium of representation: language
   - language embodies occasional cognitive innovations over many generations

   $\Rightarrow$ language as conceptual template evolving to ever higher expressive power
   - living language as “center of cognitive gravity” (24)

For Churchland, acquisition of knowledge has three levels consisting in

> *the generation of a hierarchy of prototype-representations via gradual change in the configuration of one's synaptic weights (first-level learning), and the subsequent discovery of successful redeployments of that hard-earned framework of activation-space representations, within novel domains of experience (second-level learning)...* [*and the*] *cultural assimilation of individual cognitive successes, the technological exploitation of these successes, and transmission of those acquired successes to subsequent generations, and the ever-more-sophisticated regulation of individual activities at the first two levels of learning. (22; typo corrected, emphasis deleted, order reversed)*

Let's have a closer look at the first level... (following Churchland (1989, Ch. 11))

Factors driving conceptual change: learning

Supervised vs. unsupervised learning

- supervised learning: correct output is available to agent, supervisor present
- unsupervised learning: no such correct output is available, as is commonly the case

Supervised learning

1. back-propagation procedure: desired output vector is compared element-by-element with actual output vector produced in response to training input; difference is used to compute adjustments to weights (243)
   - problem: unrealistic to assume that correct output is always available
   - no known mechanism in actual brains that does this global adjustment computation
   - poor performance for large networks
2. Boltzmann learning: hold input and output layers temporarily fixed, and repeatedly run input through network, each time adjusting weights (depending on activation level), then take next input-output vector pair, etc
   - doesn't require the computation of global error
   - still very slow for large networks
   - requires availability of correct output

Unsupervised learning

Surprising: network can learn even in absence of known output! (but they need large sample of inputs)

in unsupervised learning situations, networks must "evolve processing strategies that

(a) maximize their capacity for identifying salient information in the set of input vectors,

(b) convey such information from layer to layer in efficiently coded forms, and

(c) find similarities among the inputs so that they are taxonomized into potentially useful groupings." (246)

Hebbian learning

Characterization (Hebbian learning)

Hebbian learning is a process of weight adjustment that exploits the temporal coincidence on either side of a given synaptic junction and is therefore purely local and does not require an "outside" computation.

Hebb rules

Basic form of synaptic adjustment:

Principle (Hebbian learning)

"If a given synapse is the site of both a strong presynaptic signal and a highly activated postsynaptic cell, then the weight of that synapse in increased." (246) (Otherwise, it is decreased.)

⇒ procedure modifies weight configuration s.t. correlations among diverse elements of input signals arriving at given cell are magnified

- there exist many concrete implementations of this idea
- promise of biologically realistic procedure
- fast, no external supervisor required

So is knowledge justified, true belief?

1. It's not belief: "propositional attitude", which would require for its specification a declarative sentence, and that's unacceptable since it's too narrow (rules out non-human animals, prelinguistic children, reifies distinctions bw knowing=that and knowing-how, etc)
2. It doesn't require justification: practice of justifying epistemic commitments arises only in adult humans, almost all of our knowledge is not justified
3. It's not about truth: truth would have to be reconceptualized in order to make it applicable to sublinguistic structures (as opposed to declarative sentences)

In conclusion

- exciting new developments in science which challenge fundamental assumptions in epistemology
- philosophers cannot afford to ignore these developments
- but, against the Churchlands, I still believe more traditional epistemology to be relevant

$\Rightarrow$ important new issue: what's the relation bw cognition and learning at neural level and e.g. considerations of evidential confirmation at a higher, "linguistic" level?

- perhaps analogous to relation bw basic electronic circuitry in hardware of computer and higher programming languages...?
- Many open questions remain, and the field will not run out of work anytime soon!