Introduction to Cognitive Science: Notes

V: Neurological and Developmental Substrate

- Readings for this section: Miller *et al.* 1960, Ch.14, Some Neurophysiological Speculations; *Rizzolatti *et al.* 2002; Minsky and Papert 1988a; .

VI: Neurological and Developmental Substrate

- The ubiquitous appearance of composition and type raising in both affordance-mediated action planning of the most elementary sort on the one hand, and syntax and semantics on the other, strongly suggests that the language faculty in its syntactic aspect is directly hung onto a more primitive set of prelinguistic operations originally developed for motor planning.
- The left inferior frontal (Broca's) area that evidence from brain imaging and acquired aphasias suggests is implicated in morphosyntactic processing is immediately adjacent to areas involved in motor planning, suggesting that in evolutionary and developmental terms, the former are built upon the latter.
- The association of specific loss of *verbs* and LIF aphasias is suggestive.
- So is the fact that, when shown an object, temporal aphasic patients who cannot recover the noun ("knife") may still be able to recover its affordances ("It is to cut with", or related gesture, Miller *et al.* 1960:196).

Neural and Computational Theories

- The primate cytoarchitectonic homolog of area 44 or Broca’s area in humans, F5, has been shown by single cell recording to include “Mirror Neurons” that fire not only to specific goal oriented actions such as reaching and grasping, but also (with exquisite specificity) to the sight of another animal performing the same goal-oriented action (Rizzolatti *et al.* 2002).

- If the animal knows that the goal is not contextually valid, or if the other animals gaze is not consistent, the mere sight of motion is not enough to fire the mirror neuron.

- Other neurons in F5 fire only to the animals own actions, and/or fire to visual presentation of the object involved (Rizzolatti *et al.* 2001; Miall 2003).

- This system has usually been interpreted in terms of recognition, understanding, and imitation of the actions of other animals (Gallese *et al.* 1996).

Origins of Symbolic Representation

- It seems likely that such understanding is founded on an even more basic capability for planning the animal's own actions, of the kind proposed above.
- In particular, it seems likely that the purely motor-sensitive neurons of F5 are closely related to rules of the LDEC type, aka TOTE units or operants
- —and that the visual object-related neurons are related to the apparatus that associates objects with the actions that they afford (Miall 2003:2135).
- The interest of the mirror neurons themselves is then that their generalization over participant identities makes them necessarily **symbolic** representations, distinct from both efferent motor activity and afferent pure perception
- These units appear to map very directly onto verbs, whether we think of these as case-frames (Rizzolatti and Arbib 1998), dependency structures (Pulvermüller 2002) or CCG lexical categories discussed below.
- In CCG, such lexical items constitute the entire language-specific grammar.

A Project for a Cognitive Neurolinguistics

- This entire system is prelinguistic, rather than language-specific.
- Much of it seems to be highly localized, rather than parallel-distributed.
- However, mechanisms like Simply Recurrent Networks (SRN, Elman 1990) may well be appropriate for the process of compilation of repeated plans into compound actions and episodic memories, as opposed to novel plan construction and natural language understanding
- We need to know more about F5 in primates, specifically in relation to tool use. Are there "affordance" mirror neurons that fire both to use and appearance of tools?
- We need to understand how the planning process exploits units in F5. The limbic system seems to be implicated.

Project (Contd.)

- We need neurocomputational and machine-learning theories of how symbolic units of the kind found in F5 can be induced from sensory-motor input.
- Study of the regions adjacent to F5, (e.g. F4 which has spatially located action units Rizzolatti *et al.* 2002) and pathways to and from the cerebellum (Miall 2003) which executes and monitors them, are likely to be important.
- The computational character of the cortico-cerebellar-hippocampal sensory motor system is fairly well understood since Marr (1969)—see Gluck and Myers 2000.
- Perceptron-like reinforcement learning conditional on the intended goal state of LDEC-like operants seems to offer a mechanism for the neocortex and cerebellum and associative networks for the hippocampus.

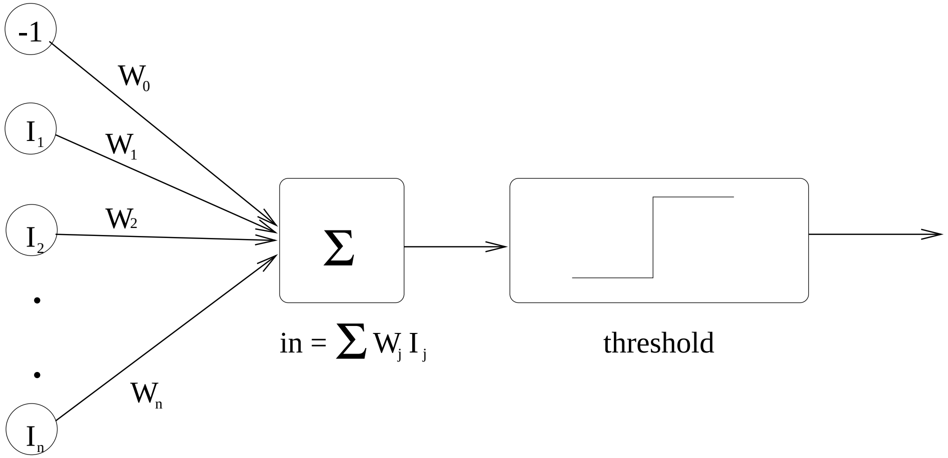
Neural Networks:Classifiers and Associative Nets

- There are two main types of general-purpose neural network computers: Classifiers, and Associative Networks.
- Both are distinguished from standard symbolic computation by “graceful degradation” under conditions of damage and noise.
- The most basic variety of classifier is the Perceptron.
- The most basic variety of associative network is the Willshaw Net.

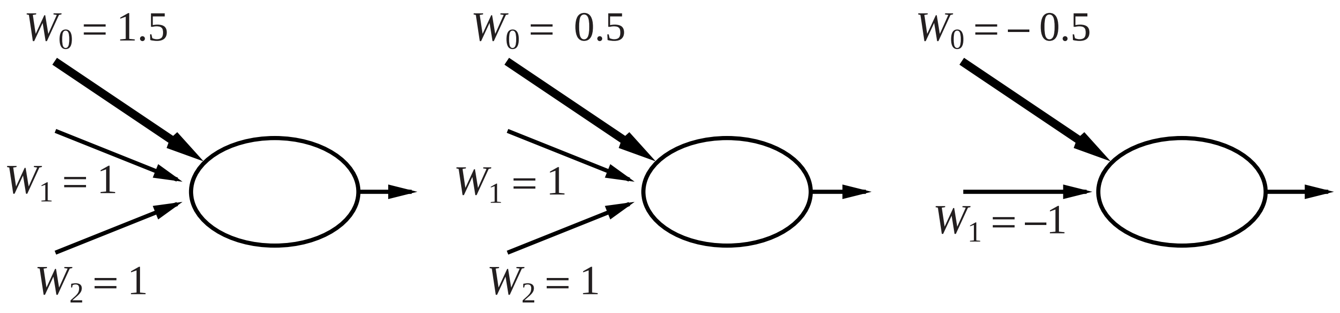
The Perceptron

A perceptron is a **single-layer feed-forward neural network**. Consider an example in which the activation function is a **step** function:

- Set $I_0 = -1$
- Unit fires when $\sum_{j=0}^{n} W_j I_j = \sum_{j=1}^{n} W_j I_j - W_0 \geq 0$
- W_0 is the *threshold*: the unit fires when $\sum_{j=1}^{n} W_j I_j \geq W_0$



Computing Boolean Functions with Perceptrons



AND **OR** **NOT**

Units with a (step) threshold activation function can act as logic gates, given appropriate input and bias weights.

AND-gate truth table

Bias	input		output
a0	a1	a2	
-1	0	0	0
-1	0	1	0
-1	1	0	0
-1	1	1	1

$$\text{AND} = \text{step}_{1.5}(1 \cdot a_1 + 1 \cdot a_2) = \text{step}_0(1.5 \cdot -1 + 1 \cdot a_1 + 1 \cdot a_2)$$

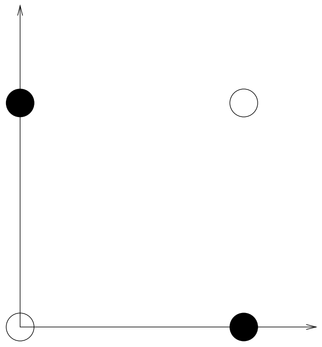
However, single-layer feed-forward nets (i.e. perceptrons) cannot represent *all* Boolean functions

Some Geometry

- In 2 dimensions $w_1x_1 + w_2x_2 - w_0 = 0$ defines a line in the plane.
- In higher dimensions $\sum_{i=1}^{n} w_i x_i - w_0 = 0$ defines a hyperplane.
- The decision boundary of a perceptron is a **hyperplane**.
- If a hyperplane can separate all outputs of one type from outputs of the other type, the problem is said to be **linearly separable**.

XOR is not linearly separable

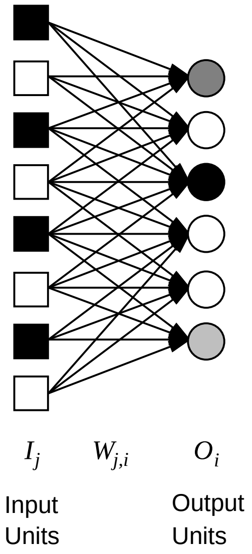
	I_1	I_2	$XOR(I_1,I_2)$
(a)	0	0	0
(b)	0	1	1
(c)	1	0	1
(d)	1	1	0



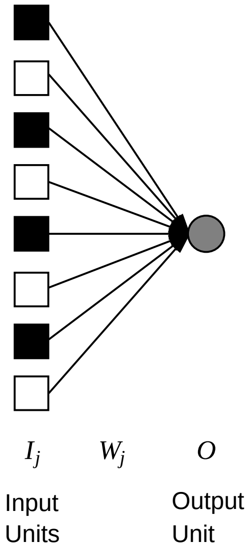
- Function as 2-dimensional plot based on values of 2 inputs
- black dot: $XOR(I_1,I_2) = 1$ and white dot: $XOR(I_1,I_2) = 0$
- Cannot draw a line that separates black dots from white ones

Single Layer — Multiple Outputs

- Each output unit is independent of the others; each weight only affects one output unit.
- We can limit our study to single-output Perceptrons.
- Use several of them to make a multi-output perceptron.



Perceptron Network



Single Perceptron

Supervised Learning in Perceptrons

- The learner sees labelled examples $e = (I_e, T_e)$ such that $f(I_e) = T_e$.
- The learner is required to find a mapping h that can be used to compute the value of f on unseen descriptions.
- In order to do that, machine learning programs normally try to find a hypothesis h that gives correct classification to the training set (or otherwise minimises the number of errors).

Learning Perceptrons — Basic Idea

- **Important Note:** We assume a threshold activation function, namely a step function, in the next few slides.
- Start by assigning arbitrary weights to $\mathbf{W}$.
- On each example $e = (\mathbf{I}, T)$:
  classify e with current network:
  $O \leftarrow \mathrm{step}_0(\mathbf{W} \cdot \mathbf{I}) = \mathrm{step}_0(\sum W_i I_i)$
  if $O = T$ (correct prediction) do nothing.
  if $O \neq T$ change $\mathbf{W}$ "in the right direction".

But what is "the right direction" ?

The Right Direction

- if $T = 1$ and $O = 0$ we want to increase $\mathbf{W} \cdot \mathbf{I} = \sum W_i I_i$
  Can do this by assigning $\mathbf{W}^{new} = \mathbf{W} + \eta \mathbf{I}$
  since $\mathbf{W}^{new} \cdot \mathbf{I} = \sum W_i^{new} I_i = \sum W_i I_i + \eta \sum I_i I_i > \sum W_i I_i$
- Amount of increase controlled by parameter $0 < \eta < 1$
- if $T = 0$ and $O = 1$ we want to decrease $\mathbf{W} \cdot \mathbf{I} = \sum W_i I_i$
  Can do this by assigning $\mathbf{W}^{new} = \mathbf{W} - \eta \mathbf{I}$
  since $\mathbf{W}^{new} \cdot \mathbf{I} = \sum W_i^{new} I_i = \sum W_i I_i - \eta \sum I_i I_i < \sum W_i I_i$
- In both cases we can assign $\mathbf{W}^{new} = \mathbf{W} + \eta \mathbf{I}(T - O)$

Perceptron Learning Algorithm (Version 1)

```
function perceptron-learning(examples) returns a perceptron hyp.
    network ← a network with randomly assigned weights
    repeat
        for each $e$ in examples do
            $O \leftarrow$ perceptron-output(network,$\mathbf{I}_e$)
            $T \leftarrow$ required output for $\mathbf{I}_e$
            update weights in network based on $\mathbf{I}_e$, $O$ and $T$
            $\mathbf{W} \leftarrow \mathbf{W} + \eta\, \mathbf{I}_e\, (T - O)$
        end
    until all examples correctly predicted or other stopping criterion
return NEURAL-NET-HYPOTHESIS(network)
```

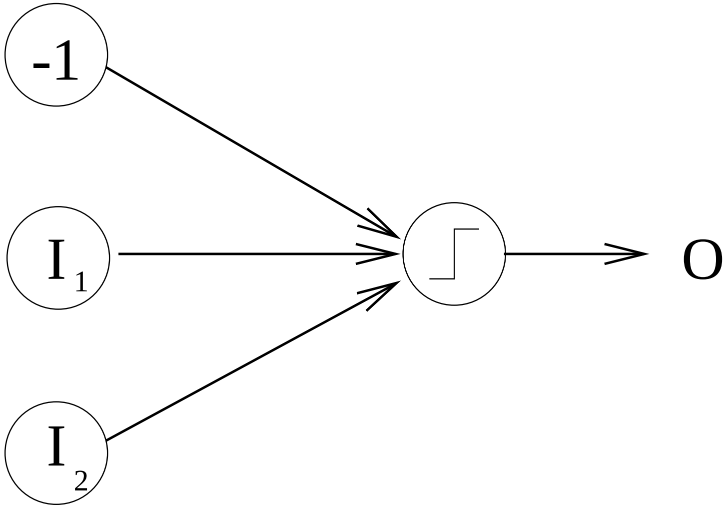
Perceptron algorithm with step (threshold) activation function.

Perceptron Learning Algorithm

- $0 < \eta < 1$ is known as the **learning rate**, other symbols e.g. α, ε used by different authors
- Rosenblatt (1960) showed that the PLA converges to $\mathbf{W}$ that classifies the examples correctly (if this is possible).
- PLA behaves well with noisy examples.
- Note that PLA given above is an *incremental* algorithm; *batch* version also possible.

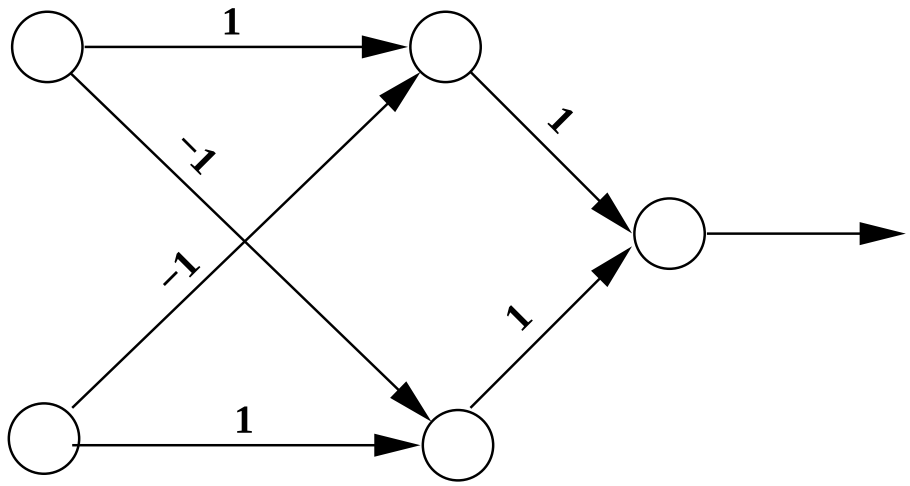
PLA:Example

- Assume that output O=1, and target is T=0, ⇒ T-O=-1
- $W_0 \leftarrow W_0 + \eta * (-1) * (-1)$
- $W_1 \leftarrow W_1 + \eta * I_1 * (-1)$
- $W_2 \leftarrow W_2 + \eta * I_2 * (-1)$



Multilayer Neural Network

- Can represent XOR using a network with two inputs, a hidden layer of two units, and one output. A step (threshold) activation function is used at each unit (threshold weights (not shown) are all zero). Many architectures possible, this is an AND-NOT OR AND-NOT network.
- In fact, any Boolean function can be represented, and any bounded continuous function can be approximated.

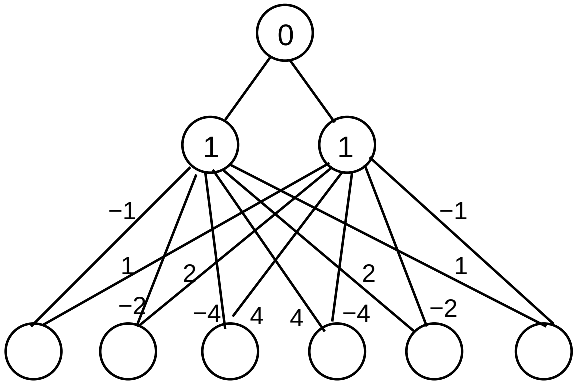


Learning Multi Layer Perceptrons

- It would be good if there were a learning algorithm for MLPs with nice convergence properties like the Perceptron Learning Algorithm.
- Rumelhart *et al.* (1986) discuss the Back-propagation Algorithm.
- However, there is no convergence theorem. BPA can get stuck in local minima.
- Minsky and Papert (1988a) argue that the MLP and BPA merely approximate the PLA in a Perceptron with exponentially growing weight values.

Learning Multi Layer Perceptrons

- For example the following MLP for detecting symmetrical six-unit input strings closely approximates the actual values learned by the BPA for Rumelhart et al.'s MLP:



- These weights appear to be growing exponentially which implies that they will take a time to learn exonential in the number of input units
- Multi layer perceptrons are interesting devices, but their training remains hard.

Willshaw Nets

- How is it that you notice a mouse behind the breadbox when all that is visible is its tail? How come you notice when someone mentions your name in the midst of a buzz of noisy conversation at a cocktail party, even when you have no idea what they said about you? What exactly is going on when you cannot remember the name of someone you are talking to, but know that it will come to you in a minute, and it does?
- These are examples of “retrieval from partial information”, “recognition from noisy input”, and “content addressable memory”. They can all be modeled in terms of massively parallel distributed processing—(M)PDP—using what are somewhat metaphorically called “Neural Networks.”
- Since even a two dimensional mouse represents rather a lot of information, we are going to look at recognition etc. of “one-dimensional” mice, represented by bit-vectors, or ordered sequences of 0s and 1s.

The Associative Net

- The Associative Net was invented by Longuet-Higgins, Buneman, and Willshaw (see Willshaw (1981)). This device illustrates three basic properties of network models which are characteristic of mechanisms involved in phenomena of human memory and attention like those mentioned above:
  - Non-localized storage (“Distributivity”)
  - Ability to recover complete stored patterns from partial or noisy input (“Graceful Degradation”).
  - Ability to work even in the face of damage (“Holographic Memory”).

The Associative Net

- An associative net acts as a distributed memory associating pairs of input and output vectors, as in the figure below, which represents a grid of horizontal input lines and vertical output lines with binary switches (triangles) at the intersections.

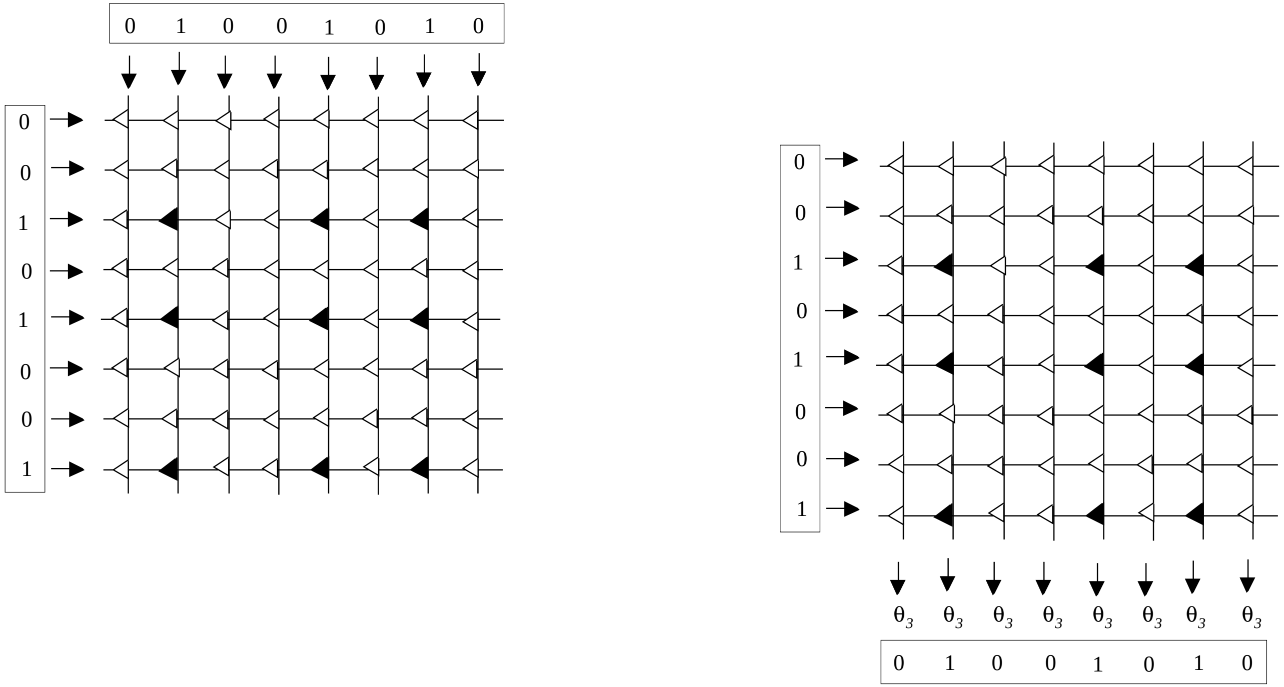


Figure 1: Hetero-associative net: Storage and Retrieval

The Associative Net

- To store an association between the input vector on the left and the output vector along the top, switches are turned on (black triangles) at the intersection of lines which correspond to a 1 in both input and output patterns.
- To retrieve the associate of the input, a signal is sent down each horizontal line corresponding to a 1 in the input. When such an input signal encounters an “on” switch, it increments the signal on the corresponding output line by one unit. These lines are then thresholded at a level corresponding to the number of on-bits in the input.

The Associative Net

- With such thresholding, an associative memory can store a number of associations in a distributed fashion, with interesting properties of noise- and damage- resistance, provided that the 1s are relatively sparse.
- For example, if one of the on-bits in the input goes off, so that we threshold at 2 rather than 3, we recover the entire associated pattern.
- Similarly if an off bit goes on we can similarly recover the correct association by reducing the threshold of 4 to 3.
- These properties depend on there being not too many similar patterns stored in the same net.

Associative net

- If patterns are "autoassociated," or stored with themselves as output, associative nets can be used to complete partial patterns, such as the mouse behind the breadbox, or someone's name from their appearance.

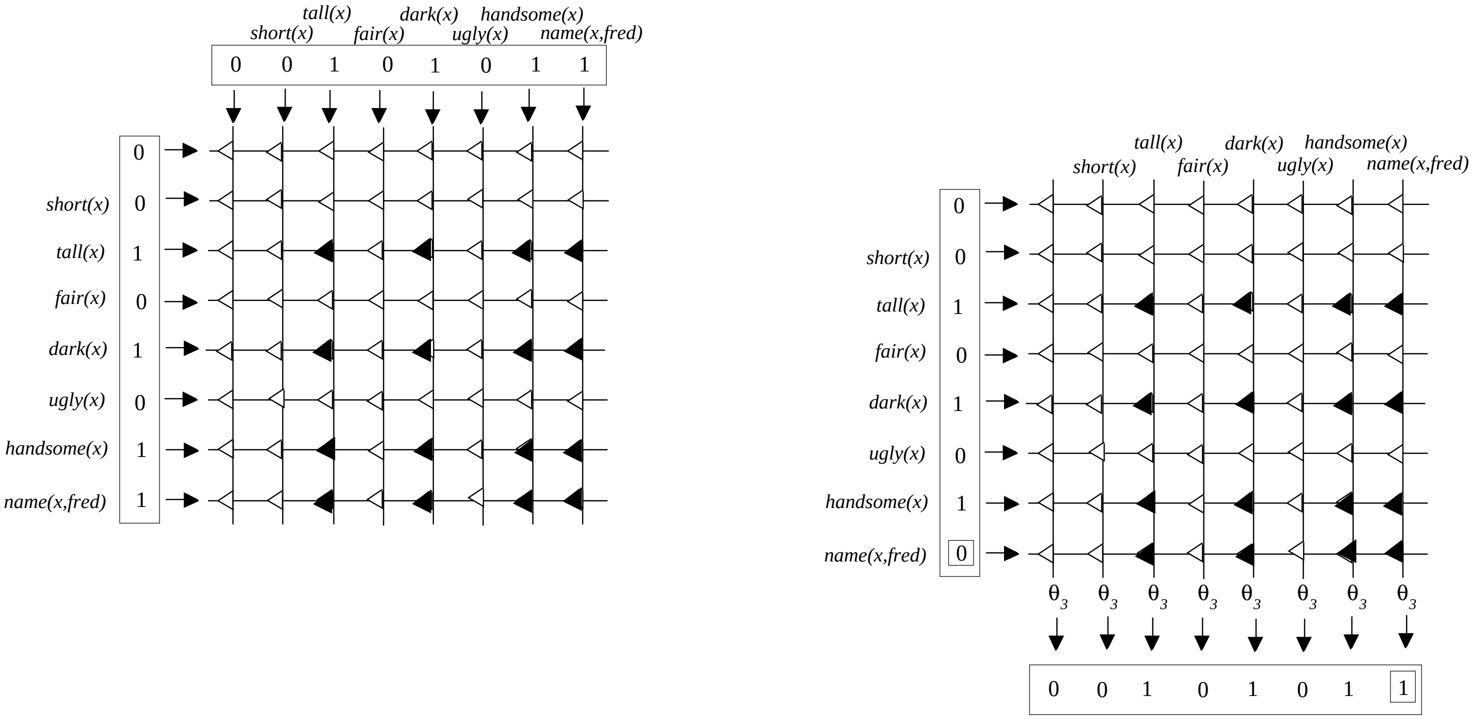


Figure 2: Auto-associative net: Storage and Augmented Retrieval

Multi-Layered Associative Networks

- The associative net can be regarded as a Multiple Output Perceptron in which the initial weights are all zero and the gain is 1.
- Just as there are multilayered perceptrons, so there are multilayered Associative Nets, such as Hopfield Nets and Recursive AutoAssociative Memory (RAAM Pollack 1990).
- Like MLP, MAN are prey to false minima and are hard to train and generalize.

Associative Networks and the Brain

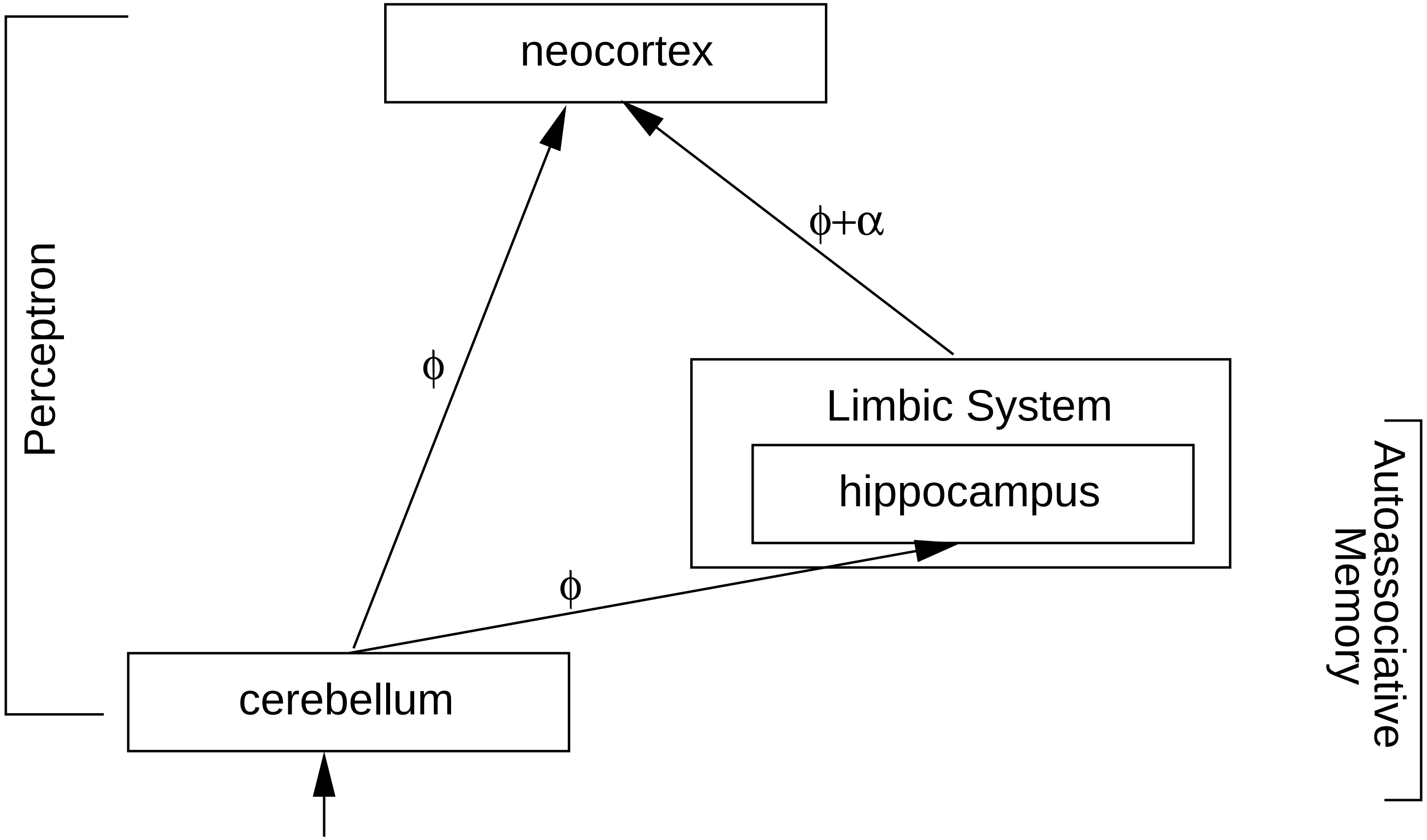


Figure 3: The basic Cerebellar-Hippocampo-Cortical dual-path circuit: (adapted from Gluck and Myers) Cf. Damasio 1999:43-47.

Associative Networks and the Brain (Contd.)

- The neural pathways to and from the motor cortex remain less clear (Daskalakis *et al.* 2004). It is likely that several levels of plan representation mediate (Wolpert *et al.* 2003).
- The process of abstracting over complete action representations needed to specify the verb/affordance-like units of F5 seems to be an open problem.
- Compositionality seems to be a general property of simple sensory motor planning.
- Plan units a.k.a. STRIPS/LDEC rules seem to be learnable with standard neurocomputational models and observable with single-cell recording.
- Abstraction, plan formation and plan execution are well understood in formal terms but remain to be understood in neurocomputational terms.

Associative Network LDEC

- Associating affordances with preconditions:

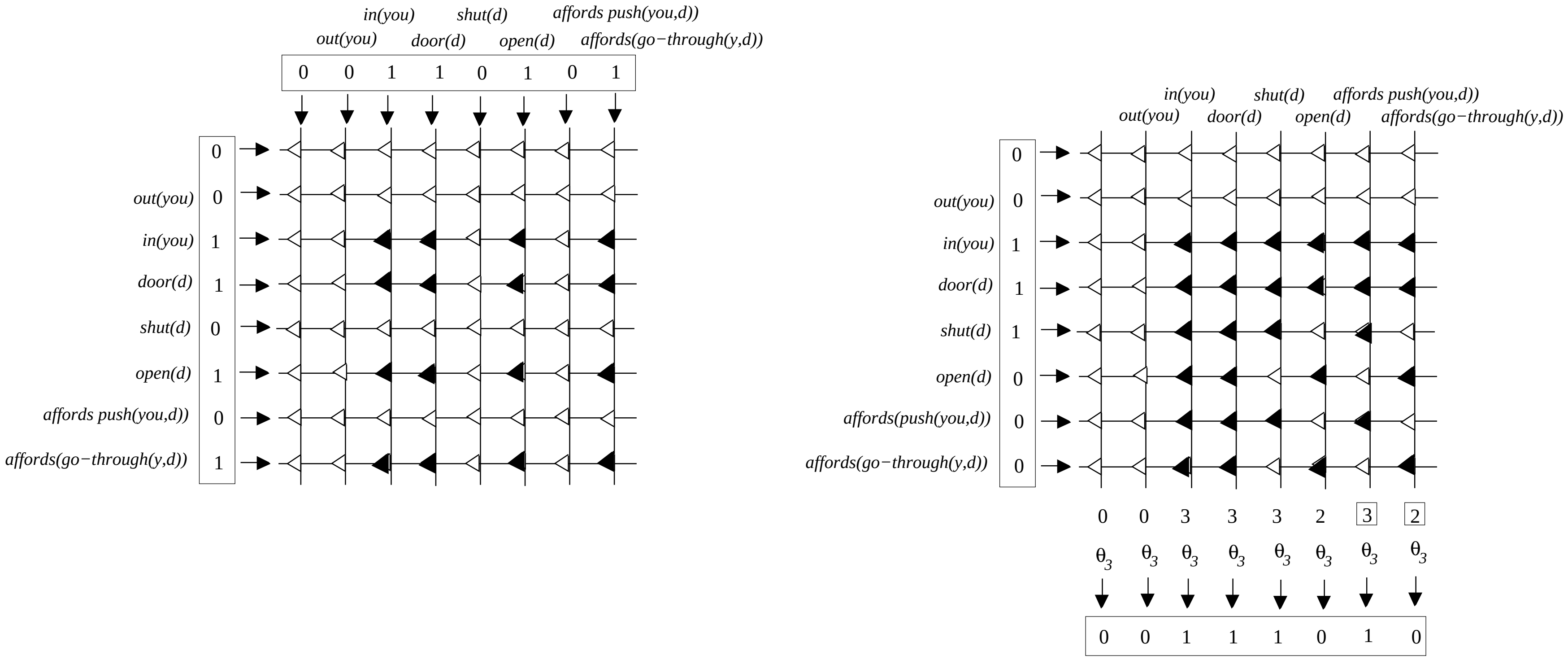


Figure 4: LDEC rules of affordance: Storing preconditions of *go-through*, and retrieving the affordance of *push* from the loaded hippocampal auto-associative net.

Associative Network LDEC

- Associating change with affordances:

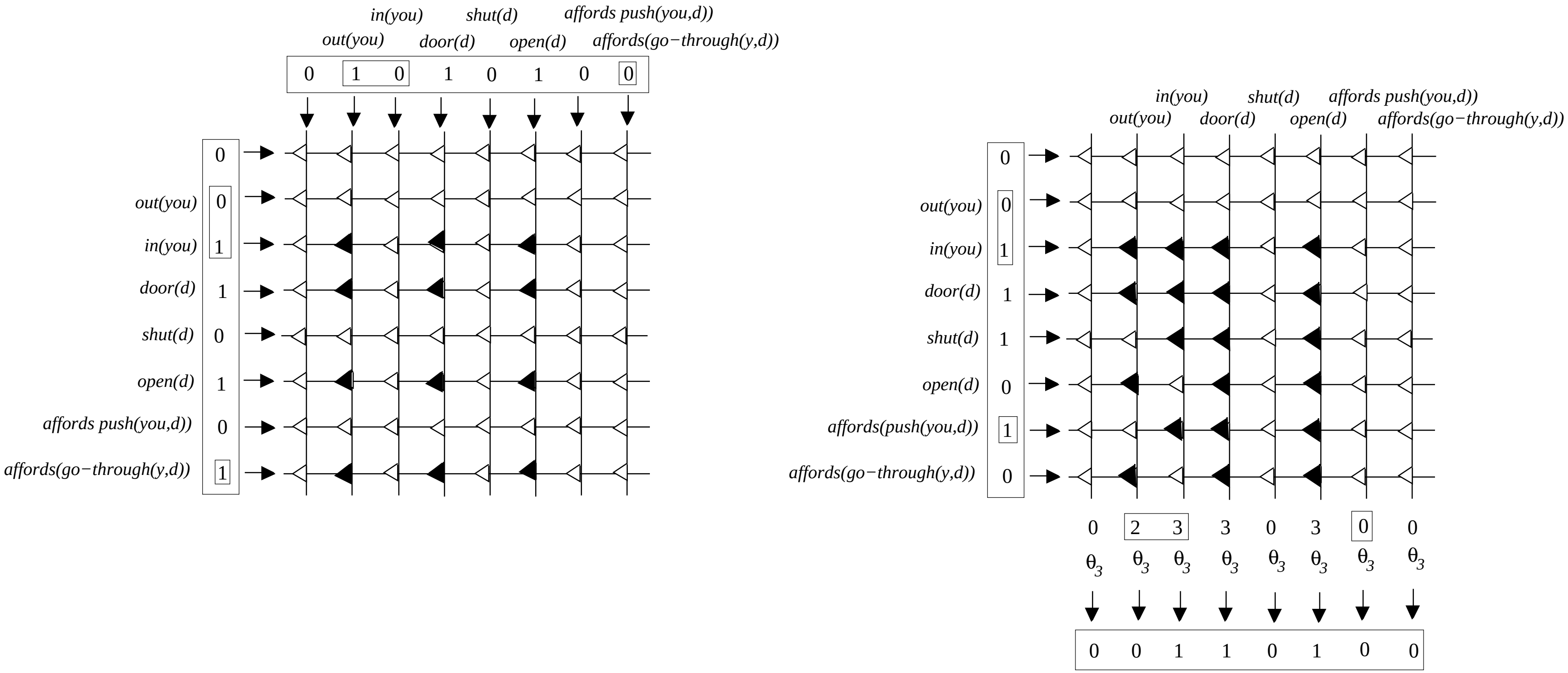


Figure 5: LDEC rules of change: Storing *go-through*, and retrieving the *push* transduction from the loaded neocortical hetero-associative net.

Associative Network LDEC

- The planning cycle:

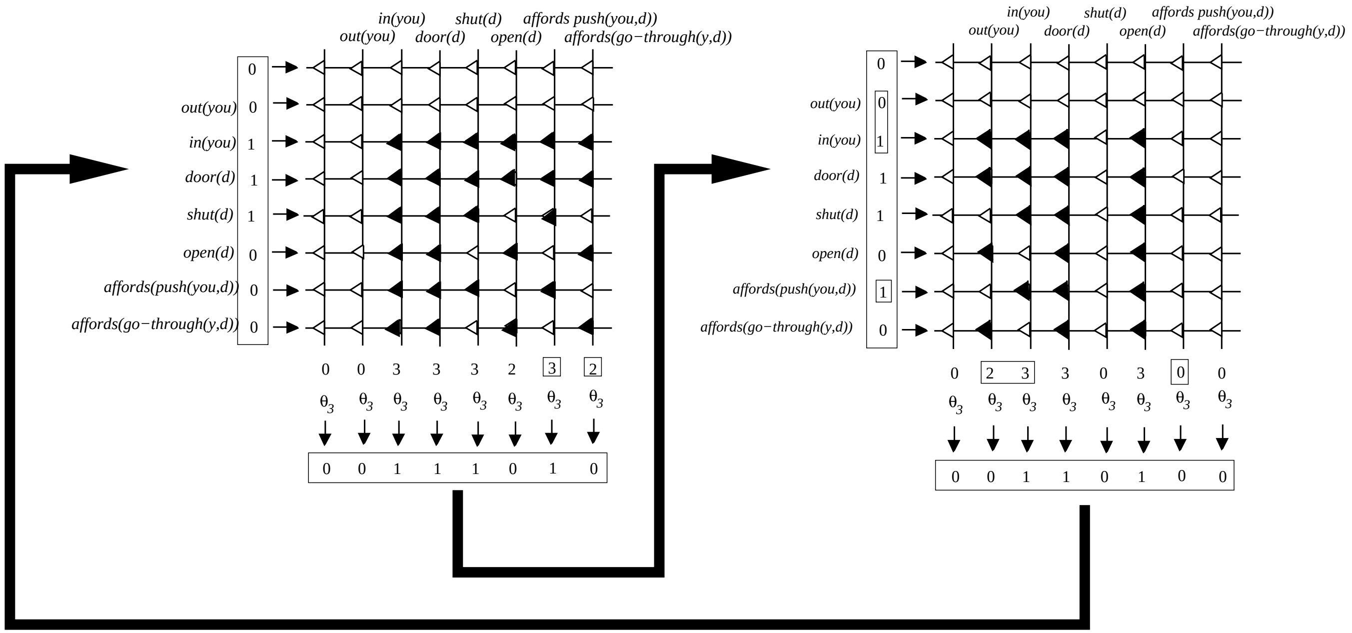


Figure 6: LDEC cycle: Retrieving the affordance of *push* from the hippocampal net, generating the next state from the neo-cortical net, and preparing to retrieve the affordance of *go-through* from the hippocampal net.

Associative Network LDEC and the Binding Problem

- These networks assume a solution tom the binding problem—that is the problem of knowing that the thing that is *open* is the same thing as the thing that is a *door*, rather than a *bottle* (or some *other* door. item One solution assumes that the input to the system is a map and that the input vector to the planning loop corresponds to a particular location having the propoerties *door*, *open*, etc.

- Some part of the input vector to the associative network cascade then represents an object/location in the map: if it is a door-location and the door-location is a shut-location and the non-object specific part of the vector says the you-location is an in-location then there is a you-push-the-door-location affordance.

Associative Network LDEC and the Binding Problem

- Object/positions in the input space suggest themselves as inputs to the planning system, via a bottom up attentional mechanism
- This general picture seems in keeping with the observations of O'Keefe and Nadel (1978), Morris *et al.* (1982), O'Keefe (1989), and McNaughton (1989), concerning single-cell recording from rat hippocampus.

Perceptron Associative Network LDEC

- The associative net can be regarded as a multiple-output Perceptron (Minsky and Papert 1969, 1988b).
- In the form it has presented so far, it is a perceptron in which the initial weights are all zero and the gain is 1.
- In order to train such a device on STRIPS rules, we had to tell it explicitly which (sparse) bits were 1s and which 0s.
- We want the machine to work that out for itself, and associate situations including things with properties like *door*, *bottle*, and *open* with actions like *push*, *go-through* and *drink*.
- The Perceptron Learning Algorithm Rosenblatt 1962, Samuel 1959, and Russell and Norvig 2003:742) will set weights on bits whose input value is irrelevant to zero.
- Applying this the the LDEC machine is work in progress

How Children Learn about Objects and Actions

- Piaget (1936), summarized by Drescher 1991 identified three major “stages” in Cognitive Development, each with several distinct substages.

- “Stages” should be though of as radically different “styles of thought,” or types of representations.

- They should not be thought of as sequential states of the child’s entire understanding, but as chronologically overlapping: a child (or an adult!) may have progressed to one stage in one domain, while remaining at a lower stage in another.

- In subsequent work, some of these distinctions have been eliminated.

The Sensory-Motor Stage

- The sensory motor stage is divided into seven substages (see Drescher 1991), whose details we ignore here, except to note that:
  - they are characterized by types (primary, secondary, etc.) of "circular reactions", which we noted earlier resemble LDEC dynamic rules, and
  - by the last of them, Sensory-Motor stage VI at around eighteen months, the child has attained a conception of the permanence of objects in the world, and is showing elementary tool use and the first distinction between the self and others.
- In many ways this is the most illuminating and enduring component of Piaget's theory.
- The main contribution of subsequent work has been to show that the developmental progression that Piaget discusses under this heading arises rather *earlier* than Piaget himself supposed.

The Operational Stage

- The operational stage is divided into two substages, the preoperational and concrete operational stages.
- **Preoperational Thought**: The child's thinking about actions and events is characterized according to Piagetian theory by:
  - An inability to take the point of view of others, and
  - To understand the constancy of physical properties like length, volume and number under reversible physical manipulations.

The Operational Stage

- **Concrete Operational Thought**: The child's thinking about concrete actions and events in terms of conservation and class inclusion is more like the adult's:
  - It reflects properties like composability, associativity, and reversibility of physical operations on entities.
  - The child can reason from the point of view of another (for example, they can lie successfully).
  - However, this thinking is rigidly tied to specific knowledge about the physical world. For example, while they can arrange three dolls in order of height, and therefore reason about concrete inequalities, they still have difficulty with abstract propositional reasoning, as with "John is taller that Bill; John is shorter than Harry; Who is the tallest?".

The Formal Operational Stage

- Formal operational thought is according to Piatgetian theory supposed to be essentially reasoning in some kind of logic, like First Order Predicate Calculus (although Piaget himself originally formulated it in rather abstruse Group-Theoretic terms.).
- The assumption is that this is the character of adult reasoning.

Experimental Work on Piaget’s Theory

- The effect of several decades of experimental work on this theory (Donaldson 1978; Wason and Johnson-Laird 1972, Stenning and van Lambalgen 2001) has been twofold.
  - **Chronology**: In almost every case except that of the transition from sensory motor stage VI to the stage of concrete operations, the experimental work has shown that the progression is earlier than Piaget thought.
  - **The structure of the Theory**: The Experimental work casts doubt on the distinction between preoperational thought and concrete operations proper (Donaldson) and on the distinction between concrete operational and formal operational thought (Johnson-Laird, Stenning).

What remains of Piaget’s Theory

- Much of the substructure of the Sensory Motor stage has stood up to experimental test, although in almost every case the succession of abilities emerges earlier.
- All other stages—including Formal Operations—collapse into Concrete Operations, in the sense that even as adults, out reasoning is tied to specifics of the world we live in (although many of these specifics are quite abstract and intangible).
- In this respect we are more reminiscent of the AI planning programs discussed in the course than of the more standard logical frameworks.

The Sensory-Motor to Operational Transition

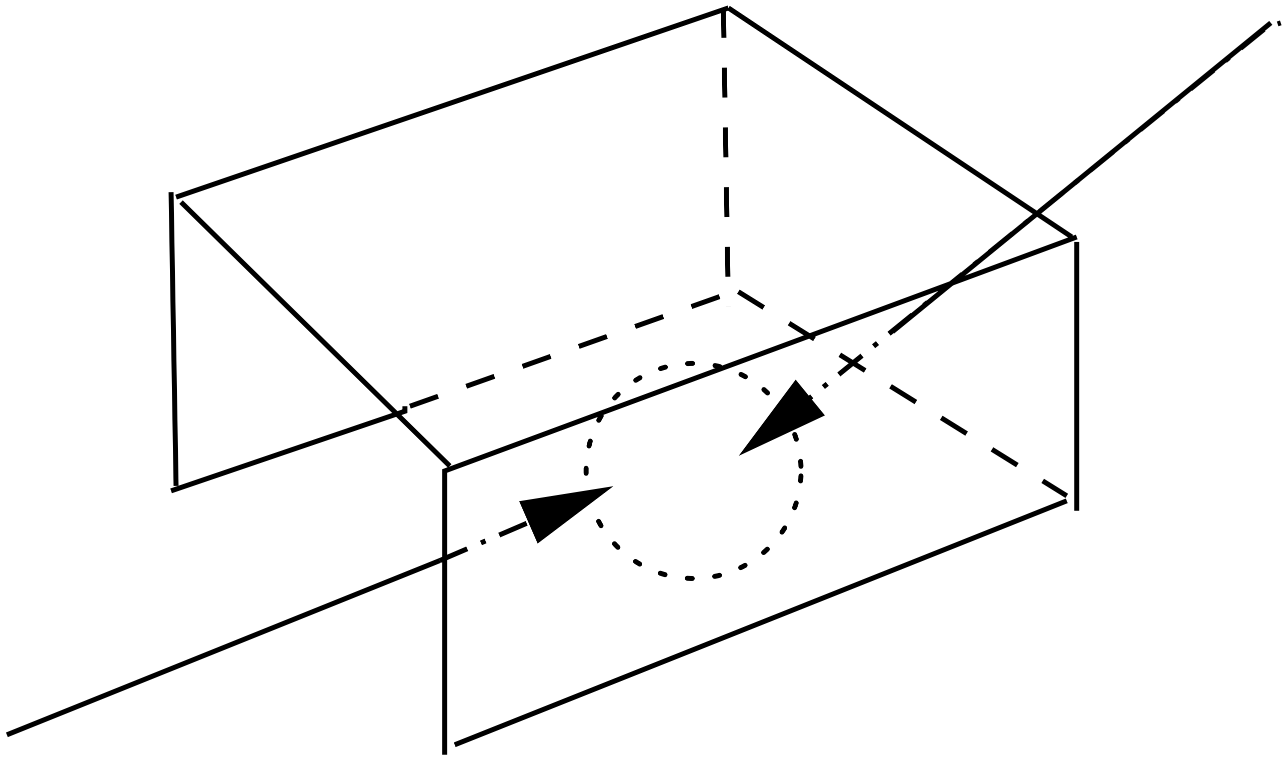
- The transition between between sensory-motor thought and what we might as well just call “Operational” thought at around eighteen months remains a dramatic and overwhelmingly important transition.
- It cannot be coincidental that this transition is closely shadowed by the onset of language.
- Does language direct this process, or merely reflect it? The explosive nature of the progress makes one believe it must be both.
- Perhaps language *not only* attaches to existing concepts, but also *directs the childs attention to concepts it would not otherwise attain*.
- Studies of blind and deaf children’s development have been extremely revealing on this point.

Plans and the Onset of Language Development

- The onset of language in infants—and the entire cognitive explosion into the Piagetian operational phase—follows closely on the mastery of motor planning involving the use of tools at the final sixth stage of the Piagetian sensory-motor phase of cognitive development.
- The onset in the child of the ability to compose motor plans such as those needed for composite reaching around an obstacle anticipates the onset of productive language use. It is also argued by Deacon (1988) and Diamond (1990) to depend on the mastery of response inhibition mediated by more frontal areas that are also implicated in language disorders.
- Damage to these more frontal areas is also characteristic of long-term impaired Broca's aphasia (Blumstein et al.)
- They are also associated with the exciting phenomenon of Mirror Neurons discussed above (Rizzolatti *et al.* 2002)

VI Plan Formation: Diamond (1990)

- The task:



to obtain the ball.

Diamond 1990 (contd., see pp.649-654)

- The box may be transparent or opaque.
- The child is free to move around.
- The box is open on one side: the child knows this.
- The child knows the object is in the box, either because it can see it through the box, or because it saw it put there, or because it put it there itself by accident.
- Can the child obtain object?

Diamond 1990 (contd., see pp.649-654)

- Until around 6.5 months, children cannot retrieve an object that is entirely inside the box.
- From 6.5 to 8 months they only succeed when they happen to get into a position where their line of sight is through the open side, or when the experimenter turns the box for them. (If the experimenter turns it back again, the child is not helped.) Then they can do a relaxation reach and obtain the object.
- From around 7.5 to 8 months the children increasingly actively change their body position, increasing their chances of the above happening.

Diamond 1990 (contd., see pp.649-654)

- From around 8.5 to 9 months, children begin to separate line of sight from line of reach and achieve seriation of reaching. In this first stage, they deliberately bend over to put themselves in a position where they can see through the open side. They then straighten up so they can no longer see along the line of reach, and do a correct two-dimensional reach.
- From 9.5 to 10.5 they progress to a second stage where they do not need to move to a position where they can look through the opening before doing a 2-D reach.
- At all stages, having a transparent box actually makes things harder, presumably because of the increased difficulty of suppressing a direct relaxation reach. (Thus the problem is not that the children forget that the object is there.)

Interim Summary

- Planning involves suppressing actions as well as seriating them.
- Planning requires a symbolic representation distinct from the motor routines that execute the plan.
- Plans are hierarchical, and partly recompiled as in Explanation-Based Learning.
- Plans are constructed by forward chaining or **B**-composition of more elementary actions.
- Plans are object-oriented, via **T**-raising of objects over their affordances.
- This makes the mapping of embedded motor routines to symbolic action representations the central problem of a neuroscience of planning.
- Hence it makes the central problem of language evolution and development the *verb*.

References

Damasio, Antonio, 1999. *The Feeling of What Happens*. New York: Harcourt.

Daskalakis, Zafiris, Paradiso, Guilermo, Christensen, Bruce, Fitzgerald, Paul, Gunraj, Carolyn, and Chen, Robert, 2004. “Exploring the Connectivity between the Cerebellum and Motor Cortex in humans.” *Journal of Physiology* 557:689–700.

Deacon, Terence, 1988. “Human Brain Evolution I: Evolution of Human Language Circuits.” In H. Jerison and I. Jerison (eds.), *Intelligence and Evolutionary Biology*, Berlin: Springer-Verlag.

Diamond, Adele, 1990. “Developmental Time Course in Human Infant and Baby Monkeys and the Neural Bases of Inhibitory Control in Reaching.” In Adele Diamond (ed.), *The Development and Neural Bases of Higher Cognitive Functions*, New York: New York Academy of Sciences. 637–676.

Donaldson, Margaret, 1978. *Children’s Minds*. Fontana.

Drescher, Gary, 1991. *Made-up Minds*. Cambridge, MA.: MIT Press.

Elman, Jeffrey, 1990. “Finding Structure in Time.” *Cognitive Science* 14:179–211.

Gallese, Vittorio, Fadiga, Luciano, Fogassi, Leonardo, and Rizzolatti, Giacomoi, 1996. “Action Recognition in the Premotor Cortex.” *Brain* 119:593–609.

Gluck, Mark and Myers, Catherine, 2000. *Gateway to Memory*. Cambridge MA: MIT Press.

Hinton, Geoffrey (ed.), 1990. *Connectionist Symbol Processing*. Cambridge, MA: MIT Press/Elsevier. Reprint of *Artificial Intelligence, 46:1-2*.

Marr, David, 1969. “A Theory of Cerebellar Cortex.” *Journal of Physiology* 202:437–470. Reprinted in Vaina 1991.

McNaughton, Bruce, 1989. “Neuronal Mechanisms for Spatial Computation and Information Storage.” In Lynn Nadel, Lynn Cooper, Peter Culicover, and Michael Harnish (eds.), *Neural Connections, Mental Computation*, Cambridge, MA: MIT Press. 285–350.

Miall, R. Christopher, 2003. “Connecting Mirror Neurons and Forward Models.” *NeuroReport* 14:2135–2137.

Miller, George, Galanter, Eugene, and Pribram, Karl, 1960. *Plans and the Structure of Behavior*. New York, NY: Henry Holt.

Minsky, Marvin and Papert, Seymour, 1969. *Perceptrons*. Cambridge, MA: MIT Press.

Minsky, Marvin and Papert, Seymour, 1988a. *Epilogue*. In Minsky and Papert (1988b), 247–280. Expanded second edition of Minsky and Papert 1969.

Minsky, Marvin and Papert, Seymour, 1988b. *Perceptrons: Expanded Edition*. Cambridge, MA: MIT Press. Expanded second edition of Minsky and Papert 1969.

Morris, Richard, Garrud, Paul, Rawlins, J, and O’Keefe, John, 1982. “Place navigation impaired in rats with hippocampal lestions.” *Nature* 297:681–683.

O’Keefe, John, 1989. “Computations the Hippocampus Might Perform.” In Lynn Nadel, Lynn Cooper, Peter Culicover, and Michael Harnish (eds.), *Neural Connections, Mental Computation*, Cambridge, MA: MIT Press. 225–284.

O’Keefe, John and Nadel, Lynn, 1978. *The Hippocampus as a Cognitive Map*. Oxford: Oxford University Press.

Piaget, Jean, 1936. *La naissance de l’intelligence chez l’enfant*. Paris: Delachaux et Niestle. translated 1953 as *The Origin of Intelligence in the Child*, Routledge and Kegan Paul.

Pollack, Jordan, 1990. “Recursive Distributed Representations.” *Artificial Intelligence* 46:77–105. Reprinted in Hinton (1990).

Pulvermüller, Friedemann, 2002. *The Neuroscience of Language*. Cambridge University Press.

Rizzolatti, Giacomo and Arbib, Michael, 1998. “Language Within Our Grasp.” *Trends in Neuroscience* 21:188–194.

Rizzolatti, Giacomo, Fogassi, Leonardo, and Gallese, Vittorio, 2001. “Neurophysiological Mechanisms Underlying the Understanding and Imitation of Action.” *Nature Reviews: Neuroscience* 2:661–670.

Rizzolatti, Giacomo, Fogassi, Leonardo, and Gallese, Vittorio, 2002. “Motor and Cognitive Functions of the ventral premotor cortex.” *Current Opinions in Neurobiology* 12:149–154.

Rosenblatt, Frank, 1962. *Principles of Neurodynamics*. New York: Spartan Books.

Rumelhart, David, Hinton, Geoffrey, and Williams, R., 1986. “Learning Internal Representations by Error Propagation.” In David Rumelhart, James McClelland, and the PDP Research Group (eds.), *Parallel Distributed Processing*, Vol. 1, *Foundations*, Cambridge, MA: MIT Press. 318–362.

Russell, Stuart and Norvig, Peter, 2003. *Artificial Intelligence: a Modern Approach*. Upper Saddle River NJ: Prentice Hall, second edition.

Samuel, Arthur, 1959. “Some Studies in Machine Learning Using the Game of Checkers.” *IBM Journal of Research and Development* 3:211–229.

Stenning, Keith and van Lambalgen, Michiel, 2001. “Semantics as a Foundation for Psychology: A Case Study of Wason’s Selection Task.” *Journal of Logic, Language, and Information* 10:273–317.

Vaina, Lucia (ed.), 1991. *From Retina to Neocortex: Selected Papers of David Marr*. Boston, MA: Birkhauser.

Wason, Peter and Johnson-Laird, Philip, 1972. *Psychology of Reasoning: Structure and Content*. Batsford.

Willshaw, David, 1981. “Holography, Association and Induction.” In Geoffrey Hinton and James Anderson (eds.), *Parallel Models of Associative Memory*, Hillsdale, NJ: Erlbaum. 83–104.

Wolpert, Daniel, Doya, Kenji, and Kawato, Mitsuo, 2003. “A Unifying Computaitonal Framework for Motor Control and Social Interaction.” *Philosophical Transactions of the Royal Society of London, B* 358:593–602.