PSYCO 452

Week 5: Value Units
Vs. Integration Devices

- Comparing the two rules
- Network by problem type interactions
- Class discussion of course to this point in the term is planned for the end of these slides

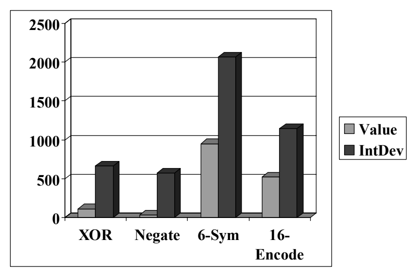
Two Rules

- We have seen that there are at least two different types of units that can be put into multilayer perceptrons
  - Integration devices
  - Value units
- Both types of multilayer perceptrons are trained with related versions of gradient descent procedures
- Is there any reason to prefer one type of network over another?

Comparing The Rules

- Early research was aimed at an empirical comparison of the two rules
- Value unit networks and comparable integration device networks were trained on a set of benchmark problems
- Results indicated interesting advantages of value unit networks

Faster Learning

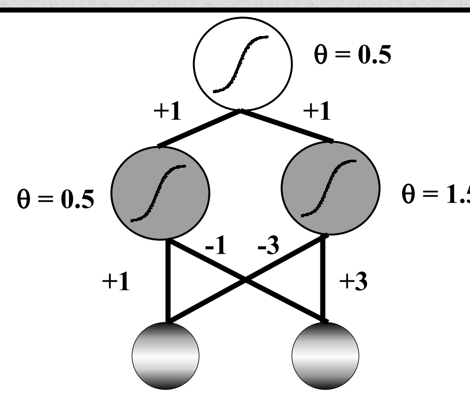


Median learning rates for a number of problems

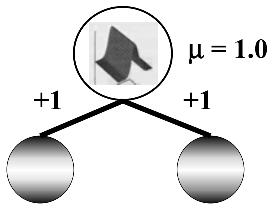
Fewer Hidden Units

- In general, value units need fewer hidden units to solve linearly nonseparable problems
- This can lead to faster learning
- This can also lead to easier network interpretation
- Consider XOR as an example:

An XOR Network



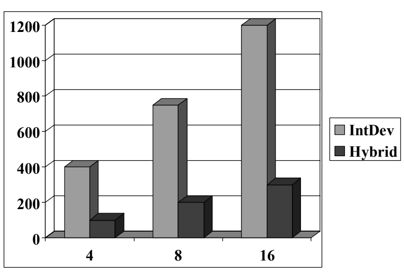
Another XOR Network



Hybrid Networks

- We can build hybrid networks, containing both types of processors
- These can be trained with the Dawson & Schopflocher rule
- This increases biological plausibility
- This also permits more controlled comparisons between networks

Hybrid Encoder Results



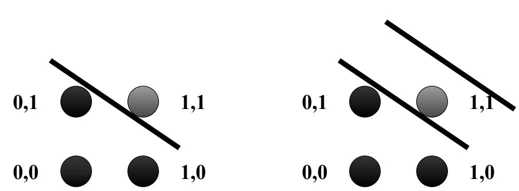
Holding Bias Constant

- Standard networks do not learn very well when bias is held constant
- Bias can be held constant in value unit networks
- This slows learning down, but leads to simpler networks
- This increases biological plausibility, because there is no evidence that neuronal thresholds are plastic

When Nets Will Be Good

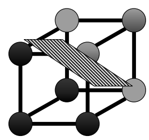
- By looking at activation functions, you can predict when networks will have problems
- Value units
  - – Good for linearly nonseparable
  - – Bad for linearly separable
- Integration devices
  - – Good for linearly separable
  - – Bad for linearly nonseparable

2 Majority Partitioning

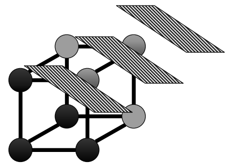


Integration Device Value Unit

3 Majority Partitioning



Integration Device



Value Unit

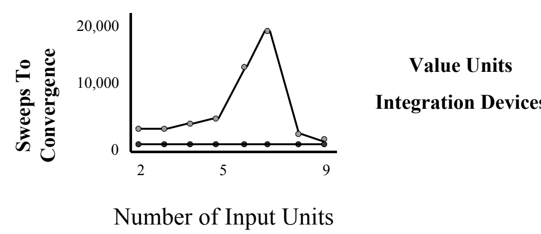
Experimental Method

- Two types of minimal networks
  - Value unit
  - Integration device
- Two types of problems (varying size)
  - majority (linearly separable)
  - parity (linearly nonseparable)
- Two dependent measures
  - learning speed
  - generalization to 25% new patterns

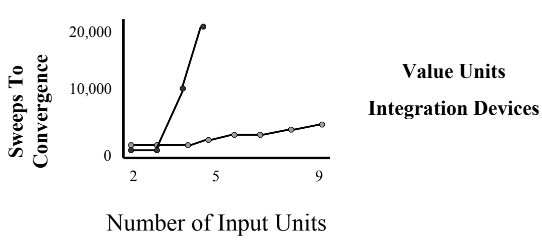
Results

- The results indicate, as predicted, a NETWORK X PROBLEM TYPE interaction
- True for both dependent measures
- This is consistent with what we know about the way that each unit carves up pattern space

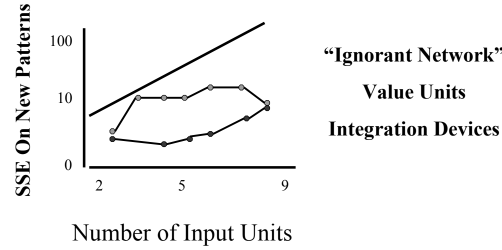
Speed For Majority



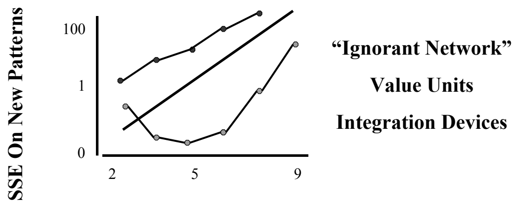
Speed For Parity



Generalization For Majority



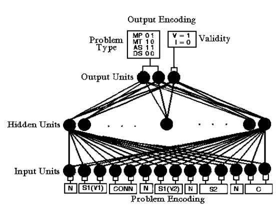
Generalization For Parity



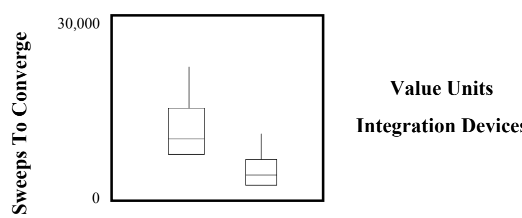
What About Other Problems

- The parity problem is extremely difficult
- Does the value unit advantage go away with simpler problems that are still not linearly separable?
- We tested this with the Bechtel and Abrahamsen (1991) logic problem

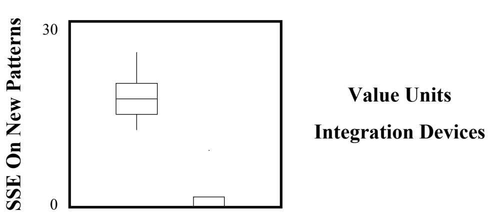
The Logic Network



Speed Results



Generalization Results



Cognitive Science?

- These results are interesting, perhaps, from a computer science perspective
- Are there other advantages of value units that might be of more interest to cognitive science?
- The answer is yes – but we will have to wait until we cover network interpretation to see why!