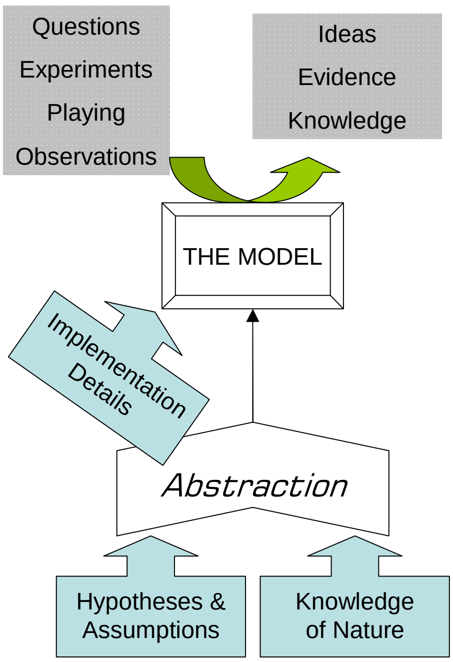
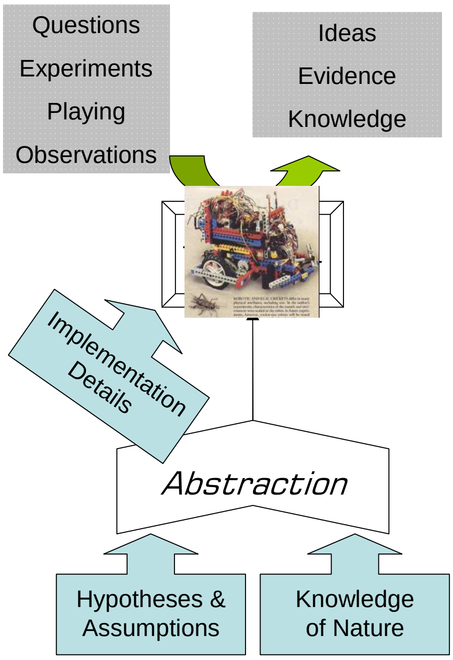
ANNs as Models

Abstractions, Levels & Questions



A model is…

- *Deliberately* simple
- *Not universal*
  - model & questions suited to each other
- Developed iteratively?
- Adjustable as part of the inquiry.
- Not necessarily improved by adding more detail!



Eg. Webb's Cricket Robot

From reading for lecture 2.

- Implementation details: lego, wheels, etc.
- Programmed with *alternative* hypotheses
- Experimented on a bit like a real animal
- Observations strengthened one hypothesis over the other

Some kinds of models

- Mathematical
- Thought experiments
- Robots
- Computer Simulations
- Animal models
  - Eg: Fruitflies for human genetics, rats for human neuroscience, and ant in the lab as a model for an ant in the woods
  - Less: humane, controllable, observable, explicit in their assumptions, precisely repeatable

Conclusions mustn't be *artefacts*

- Of the assumptions
- Of the abstractions chosen
- Of the model's implementation details

and the way the model is used.

  - Think it through
  - Vary them and check conclusions are *robust*

My diag is idealistic and not everything fits:

eg. "Look! Something as simple as how I programmed my simulation can reproduce apparently complex phenomenon X: I'm not certain X works this way, but it *may* be this simple. Or it may not!"

People often use "simulation" and "model" interchangably, and I'd rather they didn't.

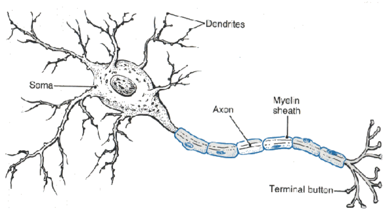
Despite changing usage, I think modelling is more in the spirit of my diagram, and simulation is more about just reproducing the appearance of something.

Some Levels of ANN (model,question)

1. Neuron-for-neuron
2. Parallel distributed system. Connectionist system.
3. A model of learning
4. A way of exploring solutions to problems

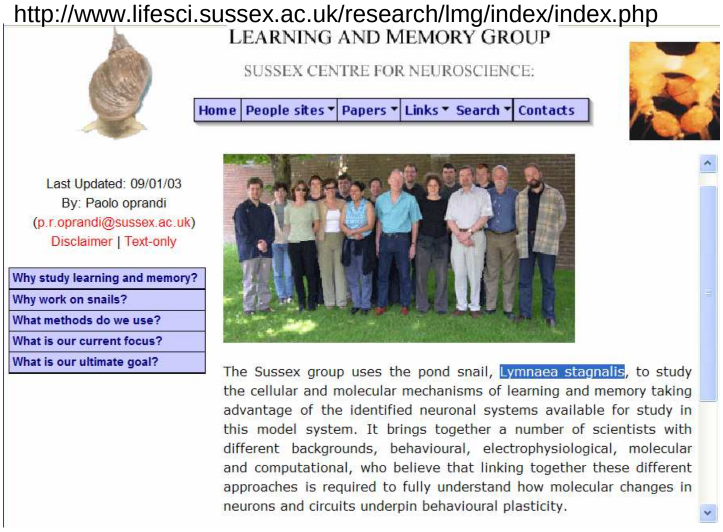
1. Neuron-for-neuron

ANN	Nature
unit	neuron
connections	synaptic interactions
weights	overall synaptic efficacies
unit's output	activity of axon (firing rate?)
transfer function	influences of weighted synapses on axon



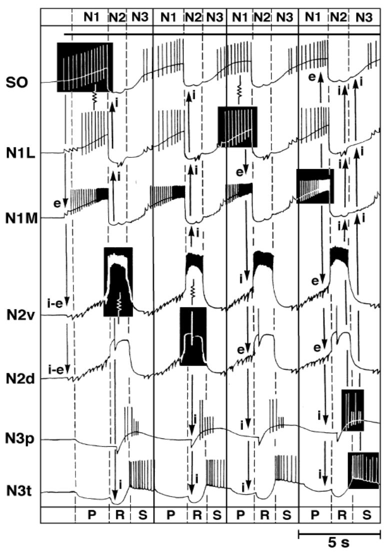
- Lots known about real neurons, but not about how ensembles work
- What details of neurons can be abstracted away when modelling ensembles?

http://www.lifesci.sussex.ac.uk/research/lmg/index/index.php



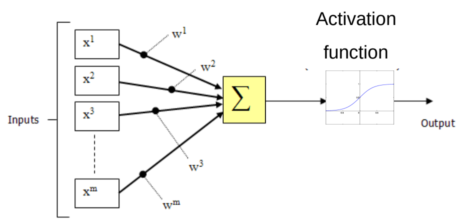
Learning in "simple" feeding network of snail as a model of learning in vertebrates etc. for cellular/molecular level questions

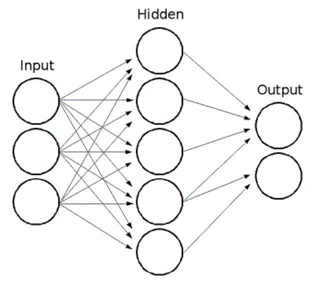
Detailed simulations



- Using the free SNNAP simulator http://snnap.uth.tmc.edu/ (Stephen Dunn, DPhil thesis)
- Even for this highly studied, small network, many unknown parameters must be twiddled to make this simulation "work"

Common worries on simple ANNs





- The only way neurons interact is through the weighted connections
- Learning is only weight change
- Traditional emphasis on primate cortex & hippocampus: regular structures, tempting to imagine work like computers
- Neuron output represented as a scalar number: assumes 'neural code' is firing rate?
- Time and dynamics neglected (cf. 'spiking' & 'recurrent dynamical' ANNs)

But!

You can't judge the adequacy of a model without knowing how it will be used.

2. Parallel Distributed System. Connectionist System

ANN	**Nature**
units	*distributed* activity
connections	interactions
weights	strengths of interactions
unit's output & transfer function	local processes

ie. whatever sort of ANN you have, it may be used as a model at a higher level than neuron-for-neuron. This is more common.

3. A model for learning

Biological learning not well understood, but we accept Hebb 1949:

"When an axon of cell A is near enough to excite cell B and repeatedly or persistently takes part in firing it, some growth process or metabolic change takes place in one or both cells such that A's efficiency, as one of the cells firing B, is increased"

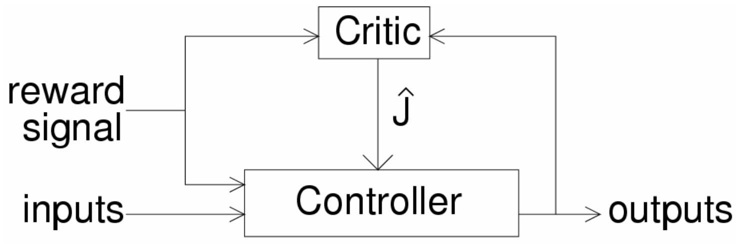
– leads to various ANN Hebbian learning rules

Remember these?

1. Supervised learning
   - During training "correct" answer for the activity of each output neuron is used. Only restricted cases imaginable in biology.
2. Reinforcement learning
   - Overall score given occasionally (eg. hunger, pain): easy to imagine in biology, hard to make work usefully in AI
3. Self-organisation
   – No training signal. Happens in nature. Hebbian learning can fit this (cf. Kohonen maps)

Internal Training Signals

- In 1&2, could some parts of the brain be the "teacher" for others?
- A tactic for reinforcement learning in AI:



Credit assignment problem: reward is delayed, so which previous actions were good and which were bad? (Eg. Inverted pendulum (cart/pole) demo)

Reinforcement Learning in AI

- Let J be a discounted sum of future rewards.
- The *critic* is trained to produce an estimate J^ of J
  - usually use the *Temporal Difference* algorithm, using the error in the difference between predictions for sucessive rewards
- This can be used by the controller to pick an action expected to give good rewards in future
  - eg. controller runs all possible outputs past the critic and picks the best one. See “Q learning”

- Evolutionary algorithms can be thought of as reinforcement learning.

ANNs to model learning at *psychological* level

- Eg. (Marchman, 1993) ANN similar to NETtalk trained to learn English grammar
- If damaged early in training, could recover through more training
- There was a “point of no return” after which damage couldn’t be repaired through further training
- Complex ideas have been suggested for why this happens in infants, but this shows the mechanisms *can be* simple.

Psychologists can use ANN models to check *biological plausibility*

- Source of training signal?
- *Locality* of signals (eg. error signals)
- Access to information (eg. no "magic sensors")
- Made of lots of little simple things? (neural)
- No central control or storage? (distributed)

- Works quickly enough, with a sensible number of units, given what we know about real neurons?

4. ANNs to explore solutions to problems, Eg:

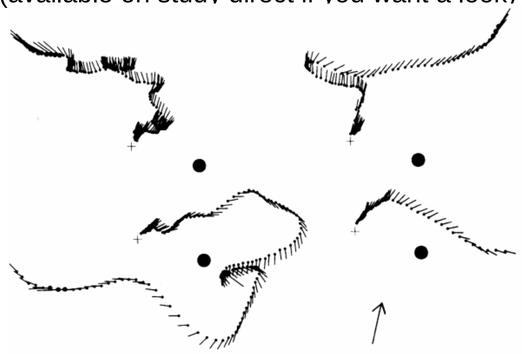
1. Train an ANN to control a robot to solve a task
2. Analyse the *behavioural strategies* rather than the ANN itself
   - *What aspects of sensory stimulation are used?*
   - *How does the robot use its own motion to manipulate the sensory world (eg. creating invariances, or active perception)*
   - *How does it use the history of interaction guide future behaviour?*

See Dave Cliff's paper (for this lecture) for what mistakes might be made!

Eg. Dale & Collett 2001

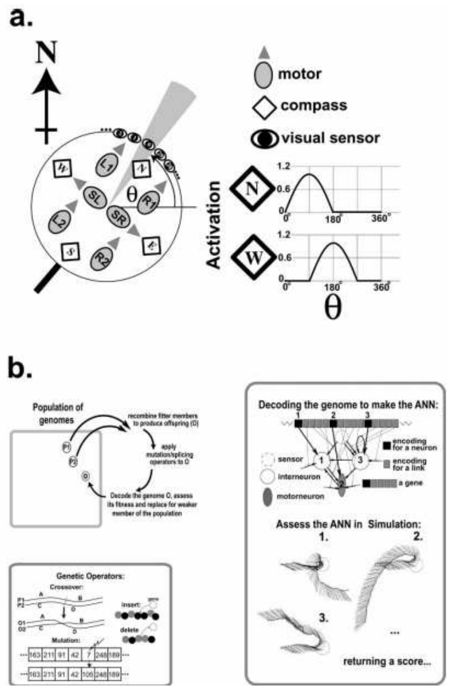
“Using artificial evolution and selection to model insect navigation”

(available on study direct if you want a look)

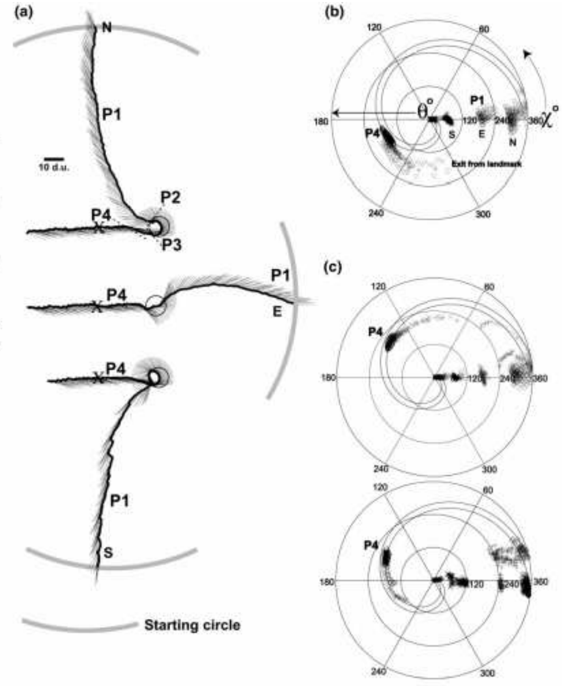


The consistency of a wasp’s approach flights to a feeder is illustrated by four approaches and landings performed by one wasp. Blobs indicate the wasp’s head, and the dash indicates the compass direction of its body axis–the same convention applies to the animats in later figures. In each approach, the wasp aims at the landmark (•) before flying to the feeder (+), where it adopts a consistent heading that remains the same throughout many approaches. Consequently, the landmark always takes up the same retinal position when the wasp is at the feeder. Data from [28].

Components of the animat and a schematic of the evolutionary algorithm that generates it. **(a)** Motors developing forward thrust are indicated by forward-pointing arrows; motors developing sideways thrust are indicated by sideways-pointing arrows. Diamonds show sensors of compass direction with sample stimulus-response curves on the right. **(b)** Evolutionary algorithm: each genome of the starting population is decoded (right) to form an ANN. The ANN is then embedded in an animat, whose fitness is assessed. Splicing the genomes of two fitter members of the population generates offspring. Genetic variation is introduced by mutation, insertion, deletion, and single-point crossover.



Evolved navigational strategies of wasplike animats. **(a)** The task and three sample trajectories. The animat is released randomly on a circle 100 du from the center of the landmark (circle) and is selected to pass as close as possible to the goal (X). The trajectory divides naturally into different phases (P1–P4). In P1, the animat fixes the left edge of the landmark with its frontal retina. P2 and P3 occur when the animat is both inside and exiting the landmark. In P4, the animat maintains a constant orientation (θ) and fixes the bottom edge of the landmark (χ) with the back of its eye. Keeping χ and θ constant at the appropriate values fixes a line between landmark and goal. **(b)** Plot of χ against θ for the three trajectories in (a). Each trajectory has a different symbol and is labeled N, E, or S according to the release point on the starting circle. The two spirals show values of χ and θ that will maintain a line to the goal for the two edges of the landmark. **(c)** χ-θ plots for two further wasplike animats show that all three animats have evolved the same basic strategies.



Artificial Evolution

AKA ‘evolutionary computing’, ‘evolutionary algorithms’, ‘genetic algorithms’ etc. Takes basic idea of selection acting on heritable variation and uses it to design things for us. Applied to ANNs, we can evolve arbitrary architectures: recurrent, dynamical, closer to biology, etc.