Overview – introducing three kinds of internal mental representation that are invoked by Cognitive Scientists to explain human behaviour.

Neural Networks

Production Systems

Logic

These representations are used in different ways by Cognitive Scientists – we will looking at how they are used

But first, we will consider the human brain
Biological Intelligence in humans

Biological intelligence in humans is derived from a biological organ: the brain

What kind of brain organisation can you see in this image?
There is a big distinction between the cortex and the rest of the brain. Humans have a very large cortex compared with non-cortical brain areas. Which areas are in the cortex and which are sub-cortical?
Are these diagrams using the same scale?

The cortex regions are coloured green.

What are the relative sizes of cortical tissue in humans and rats?
Two views showing the outside cortical lobes and inner limbic system
Brain structures are composed of layers of neurons, and each neuron has component parts.
Brain structures and cognitive modelling

Different brain structures can be composed of different kinds of neurons, and these neurons can also be organised differently (for example different numbers of layers or types of connection).

Different brain structures also give rise to different kinds of behaviours. Lots of different kinds of cognitive models draw upon the particular anatomy and function of parts of the brain

Hippocampus – forming memories
Amygdala – involved in emotional responses (fight or flight)
Basal Ganglia – action selection and movement
Frontal lobe – setting goals and intentions and planning
Temporal lobe – language
Occipital lobe – vision
Parietal lobe – movement

How does the way the brain processes information compare with a computer?
How to simulate the actions of neurons

ANN’s and Production Systems both simulate the actions of Neurons, but in very different ways.

However, what they have in common is that they can both be viewed as signal detectors.

Cognitive Scientists also use logic to model human intelligence, But how do simulations based upon logic relate to ANN and Production System simulations?

To explain this we will look at the work of Allen Newell, in particular his influential book: *Unified Theories of Cognition*
What is a symbol system?
Digital computers and Turing machines, carry out computations by the manipulation of symbols according to syntactic rules.

What is a knowledge system?
Rational agent, using rational rules to guide behaviour

Newell (1990): “humans are symbol systems that are at least modest approximations of knowledge systems. They might be other kinds of systems as well, but at least they are symbol systems”
Evidence that humans are symbols systems

Weak evidence:
humans can approximate Universal Turing Machines

Strong evidence:
flexibility of human cognition

humans can reason about problems, and reason about what affect actions will have before they take those actions. In the words of Karl Popper: “Good tests kill flawed theories: we remain alive to guess again”

Using the ability to represent situations, and reason about situations, that you have never experienced (and humans during evolution didn't experience either)
Digression: How are Symbol systems and Knowledge systems implemented in artifacts (ie computers)?

How are artificial symbol systems, like the java programming language implemented in the physical universe?

What is an example of an artificial knowledge system?

How are artificial knowledge systems implemented in the physical universe?

**Draw a tower of levels of implementation from atoms through intermediate levels to an artificial knowledge system**

What would happen if we interrupted the processing of this systems after: a milli-second; a second; an hour?
Newell conjectured that humans are formed of knowledge systems and symbol systems that are built up in a similar way to computers.

<table>
<thead>
<tr>
<th>Band</th>
<th>Time Scale</th>
<th>Representation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Social band</td>
<td>$10^5$-$10^7$ seconds</td>
<td>(months, weeks, days)</td>
</tr>
<tr>
<td>Rational band</td>
<td>minutes, hours</td>
<td>(what representation?)</td>
</tr>
<tr>
<td>tasks</td>
<td>$10^4$-$10^2$ seconds</td>
<td></td>
</tr>
<tr>
<td>Cognitive band</td>
<td>(100ms-10sec)</td>
<td>(what representation?)</td>
</tr>
<tr>
<td>units tasks</td>
<td>10 seconds</td>
<td></td>
</tr>
<tr>
<td>operations</td>
<td>1 seconds</td>
<td></td>
</tr>
<tr>
<td>deliberate acts</td>
<td>$10^{-1}$ seconds</td>
<td></td>
</tr>
<tr>
<td>Biological band</td>
<td>(100μs – 10ms)</td>
<td>(what representation?)</td>
</tr>
<tr>
<td>neural circuits</td>
<td>$10^{-2}$ seconds</td>
<td></td>
</tr>
<tr>
<td>neurons</td>
<td>$10^{-3}$ seconds</td>
<td></td>
</tr>
<tr>
<td>organelles</td>
<td>$10^{-4}$ seconds</td>
<td></td>
</tr>
</tbody>
</table>

Note: most cognitive task take less than 10 seconds. If they are allowed to complete they form part of rational decisions. What if they are interrupted and not allowed to complete? (see role of amygdala in previous slides)
Newell views humans (and computers) as nearly decomposable systems

Simon (1996): “In hierarchichal systems we can distinguish between the interactions among subsystems, on the one hand, and the interaction within subsystems – that is among the parts of those subsystems – on the other ... we may move to a theory of nearly decomposable systems, in which the interactions among the subsystems are weak but not negligible”

What are the hierarchical subsystems in computers and humans?

Newell views the levels for humans on the previous slide as nearly decomposable, that is each level is built upon the previous level with only weak interaction between levels
How does logical reasoning fit into Newell's framework?

Newell notes that cognitive processes must complete at some point. When they do they are equivalent to the process of logical reasoning.

So where consideration of timing provides a graduated set of constraints for ANNs and Production Systems there is a single constraint for simulations that are based on logical reasoning, and that is that the task has time to run to completion.
Two main problems with Newell's ideas about humans as symbol systems and knowledge systems

1. Information processing between levels in humans is not 'nearly decomposable'. Lower level processing can have significant effects in higher level processing, so that humans are not truly rational (even if they have enough time to be rational).

2. Lots of Cognitive Scientists do not invoke Neural Networks just as low level representations, working at short time-scales, of the type that Newell described (though many neuroscientists obviously do).

When Cognitive Scientists invoke Neural Networks to explain higher level cognitive tasks, they are not considering these representations in an implementational capacity, but at an abstract algorithmic level. High level Neural Network models can be much more liberal in the learning rules they possess.
Both Production Systems and Neural Networks can be considered as implementational levels and as abstract information processing algorithms.

Neural Networks can work like actual neurons, or have biologically implausible learning rules.

Production systems can run arbitrary procedures in the programming language they are implemented in (ie pop11, lisp, java), or they can be constrained to fire with the same temporal properties as the Basal Ganglia (upon which many are based).
PDP connectionist models are not at the implementational level

Rumelhart and McClelland (PDP, ANNs):

work not situated just at implementational level

had very good successes at explaining linguistic phenomena:

  grammar formation
  past tense formation

Recent work on PDP++ and the leabra algorithm (O'Reilly 2006)
(more lower level constraints – but also working towards computational universality)
ACT-R (adaptive control of thought – rational) is an important Production System model that is constrained by timing information and data from imaging studies.

ACT-R is a leading cognitive architecture, it supports a number of subsystems with their own representations within a single architecture.

It explains (predicts) a lot of human behaviour, in experiments, in naturalistic settings such as using cockpits or computers.

Its operation can be seen in imaging experiments.
In between each system are buffers that hold information for a set amount of time, and then let it decay, like forgetting. So the buffers are like short-term memory. We can speculate that the contents of buffers are the mental contents that a human is conscious of.
Evidence of a bias from completely rational abstract reasoning: Which cards do you turn over to test if this rule is true?

'If there is a vowel on one side of a card, then there is an even number on the other side'
'If you borrow my car, then you have fill up the tank with petrol'

Borrowed car
Did not borrow car
Filled up tank with petrol
Did not fill up tank with petrol
Newell's assumption of near decomposability between system levels is also suspect.

Tooby and Cosmides (2004) provide several examples of how humans possess cognitive instincts (that is special purpose evolved mental modules) that are adaptations for social living.

The implication of this is that humans do not act as similarly to knowledge level systems as other researchers might like to believe.
Conclusion

Neural Networks, Production Systems and Logic are all invoked by Cognitive Scientists, sometimes in different ways. There is more than one way to compare these modelling frameworks (such as levels of description).

Newell provides one way to view these representations, but has some problems. However, Newell's view does capture the difference between logic and the other representations.

Humans are not as rational as some would like to think. They do possess rationality, its just that it is an evolved form of rationality. That is, it is rationality that has been adaptive in the evolutionary past of the human species.