AITA : Brain Modelling

1. Brain Modelling – What Needs Modelling?
    - Development – Learning and Maturation
    - Adult Performance Measures
    - Brain Damage / Neuropsychological Deficits
2. Validating the Models – Experimental Testing
3. Modelling Simple Cognitive Tasks – A Case Study
    - Models of Reading Aloud / Lexical Decision
4. Modelling More Complex Human Abilities
5. Implications for Building AI Systems

Brain Modelling – What Needs Modelling?

It makes sense to use all available information to constrain our theories/models of real brain processes. This involves gathering as much empirical evidence about brains as we can (e.g. by carrying out psychological experiments) and comparing it with our models.

The comparisons fall into three broad categories:

Development : Comparisons of children's development with that of our models – this will generally involve both maturation and learning.

Adult Performance : Comparisons of our mature models with normal adult performance – exactly what is compared depends on what we are modelling.

Brain Damage / Neuropsychological Deficits : Often performance deficits, e.g. due to brain damage, tell us more about normal brain operation than normal performance.

We shall first look at the general modelling/testing issues involved for each of these three categories, and then consider some typical experimental and modelling results in more detail for a particular case study: reading aloud and lexical decision.

Development

Children are born with certain innate factors built into their brains (e.g. it already has a modular structure). They then learn from their environment (e.g. they acquire language and motor skills). Many systems also have maturational factors which are largely independent of their learning environment (e.g. they grow in size). Some children have developmental problems (e.g. dyslexia, strabismus).

Psychologists spend considerable effort in studying these things. Typically they measure the order in which various skills are acquired (and sometimes lost), the ages at which particular performance levels are reached, and they also try to identify pre-cursors to abnormal development.

It is often difficult to tell which abilities are innate and which are learned (the Nature-Nurture debate). Compensatory strategies can make it difficult to identify the causes of developmental problems. Ethical restrictions often make empirical studies difficult.

We aim to build models (e.g. involving neural networks) that match the development of children. These models can then be manipulated in ways that would be unethical with children, or simply impossible to carry out in practice.

Adult Performance

If we succeed in building accurate models of children's development, one might think our adult models (e.g. fully trained neural networks) will require little further testing. In fact, largely due to better availability and reliability, there are a range of adult performance measures that prove useful for further constraining our models, such as:

Accuracy : basic task performance levels, e.g. how well are particular aspects of a language spoken/understood, or how well can we estimate a distance?

Generalization : e.g. how well can we pronounce a word we have never seen before (vown fi gowpit?), or recognise an object from an unseen direction?

Reaction Times : response speeds and their differences, e.g. can we recognise one word type faster than another, or respond to one colour faster than another?

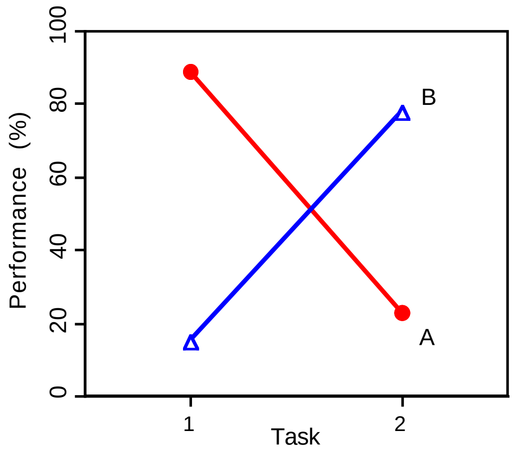
Priming : e.g. if asked whether *dog* and *cat* are real words, you tend to say yes to *cat* faster than if you were asked about *dot* and *cat* (this is *lexical decision priming*).

Speed-Accuracy Trade-off : across a wide range of tasks your accuracy tends to reduce as you try to speed up your response, and vice-versa.

Different performance measures will be appropriate to test different models. For brain modelling, the more human-like the models the better. Often we try, and sometimes succeed, in building AI systems that perform better than humans.

Brain Damage and Neuropsychology

Most tasks can be accomplished in more than one manner. For example, there are many cues that might be used to focus our eyes for objects at different distances, and it can be hard to determine how humans actually use those cues. Often, the errors produced by brain damaged patients provide valuable evidence of mental structure (Shallice, 1988). The inference from Double Dissociation to Modularity is particularly important:



Double Dissociation

If Patient A performs Task 1 well but is very poor at Task 2, and Patient B performs Task 2 well but is very poor at Task 1, we say that there is a *Double Dissociation*. From this we can usually infer that there are separate modules for the two tasks.

The detailed degradation of performance due to different types of brain damage can be used to infer how normal performance is achieved. Naturally, if our brain models do not exhibit the same deficits as real brains, they are in need of revision (Bullinaria, 2002).

Validating the Models – Experimental Testing

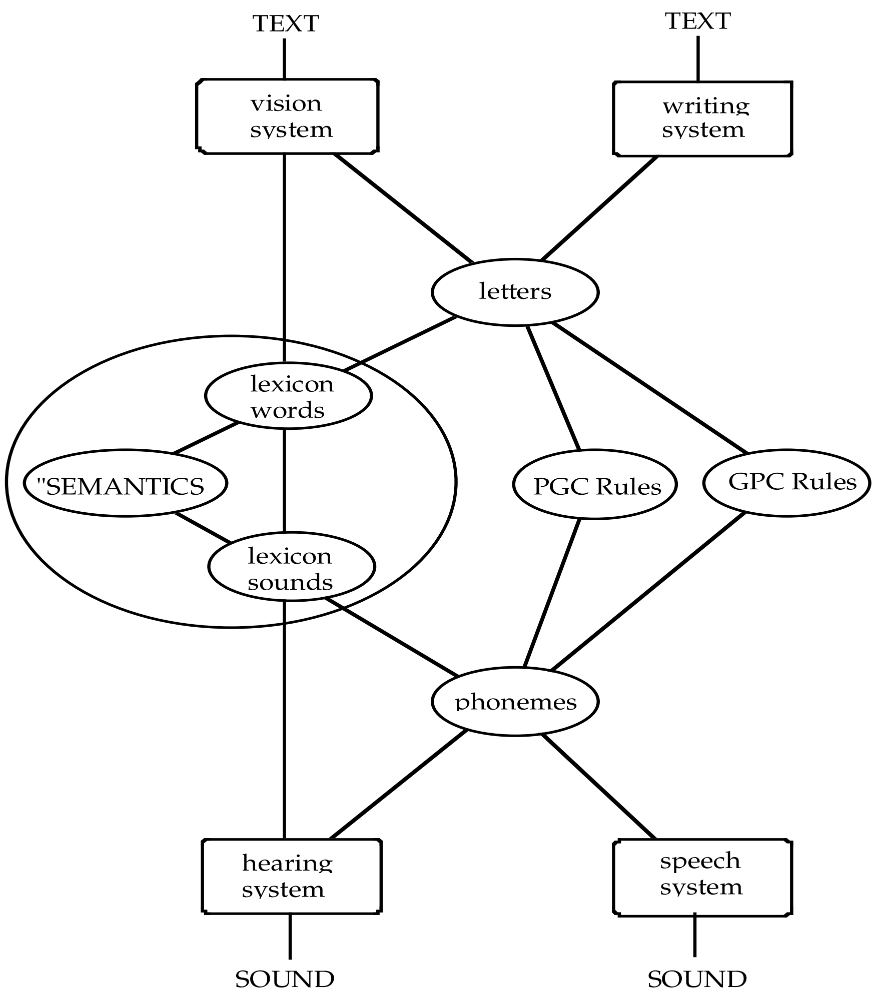
Psychologists have devised numerous ‘ingenious’ experiments to test human abilities on a range of tasks, and hence constrain our models of how we carry out those tasks. To illustrate this, we shall concentrate here on two particularly simple tasks:

Naming / Reading Aloud Present the experimental subject with a string of letters and time how long it takes them to read the word aloud. Count and classify the errors. The letter strings may be words of different frequency and regularity, or they may be pronounceable made-up words (non-words). This should give clues on how the mappings between *graphemes* (letters) and *phonemes* (sounds) are organised.

Lexical decision Present the experimental subject with a string of letters (or sounds) and time how long it takes them to decide whether it is a real word or a non-word. See if changing the preceding string makes a difference (i.e. priming). This gives clues on how the mappings between graphemes (letters) or phonemes (sounds) and the ‘lexicon’ or ‘store of word meanings’ are organised.

It turns out that some very simple neural network models can account for a surprising range of experimental data (Plaut & Shallice, 1993; Plaut et al., 1996; Bullinaria, 1997).

Traditional Dual Route Model of Reading & Related Tasks



Traditionally tasks such as reading were modelled in terms of “boxes and arrows” with each box representing a particular process (e.g. a set of rules for converting graphemes to phonemes), and arrows representing the flow of information. One then modelled brain damage by removing particular boxes or arrows. This actually accounts for a lot of human empirical data (Coltheart et al., 1993). However, recent neural network models (Plaut et al., 1996; Bullinaria, 1997) have been able to simulate much finer grained empirical data. We shall look in turn at a number of the relevant modelling issues and empirical results.

Representation Problems for Reading Aloud

To set up a neural network reading model we must first decide on appropriate input and output representations. There are three basic problems that must be addressed:

Alignment Problem The mapping between Letters and Phonemes is often many-to-one:

e.g. ‘th’ $\rightarrow$ /D/ and ‘ough’ $\rightarrow$ /O/ in ‘though’ $\rightarrow$ /DO/

It is not obvious to a network how the Letters and Phonemes should line up.

Recognition Problem The same letters in different word positions and different words should be recognized as being the same:

e.g. ‘d’ in ‘deed’ $\rightarrow$ /dEd/ and ‘fold’ $\rightarrow$ /fOld/

Context Problem The same letters in the same positions in different words are often pronounced differently :

e.g. ‘c’ in ‘cat’ $\rightarrow$ /kat/ and ‘cent’ $\rightarrow$ /sent/

We have a complicated hierarchy of rules, sub-rules and exceptions. Fortunately, neural networks are very good at learning such things.

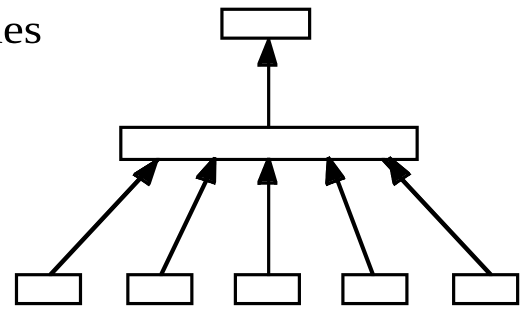
The Multi-target *NETtalk* Model

The ***NETtalk model*** of Sejnowsi & Rosenberg takes care of the recognition and context problems. Each output phoneme simply corresponds to letter in middle of input window:

output - phonemes (nphonemes)

hidden layer (nhidden)

input - letters (nchar • nletters)



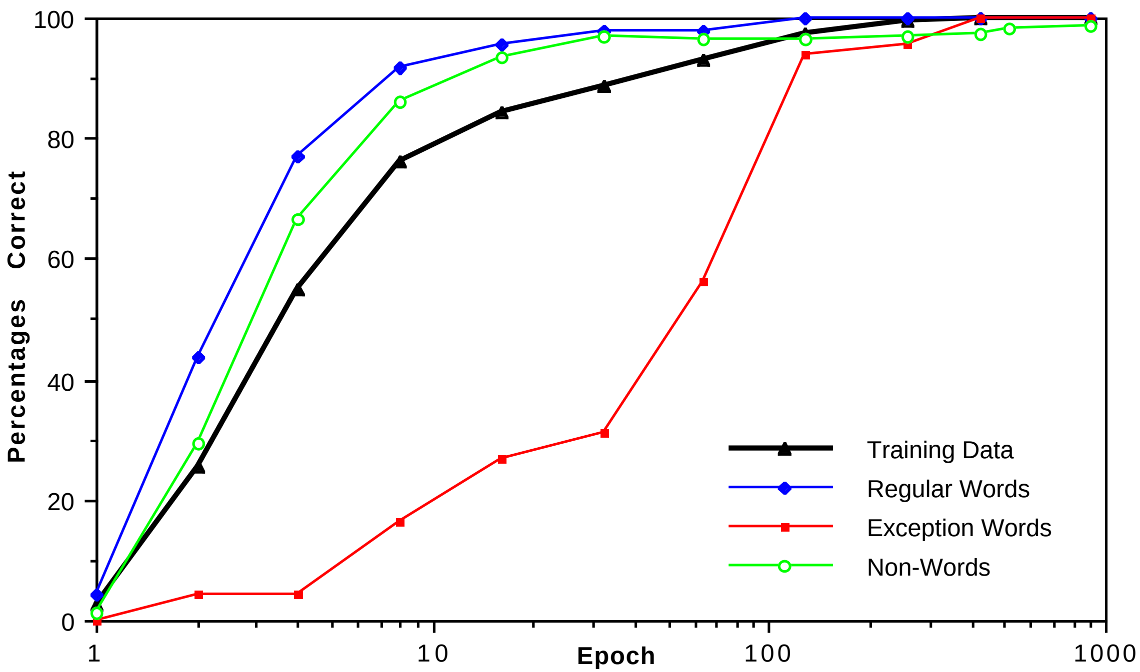
It turns out that the network can also figure out the alignment problem by assuming the alignment that best fits in with its expectations (Bullinaria, 1997), e.g. for ‘ace’:

presentation	inputs	target outputs
1.	- - - **a** c e -	A A -
2.	- - a **c** e - -	s - A
3.	- a c **e** - - -	- s s

It can then be trained with a standard learning algorithm (e.g. back-propagation).

Development = Network Learning

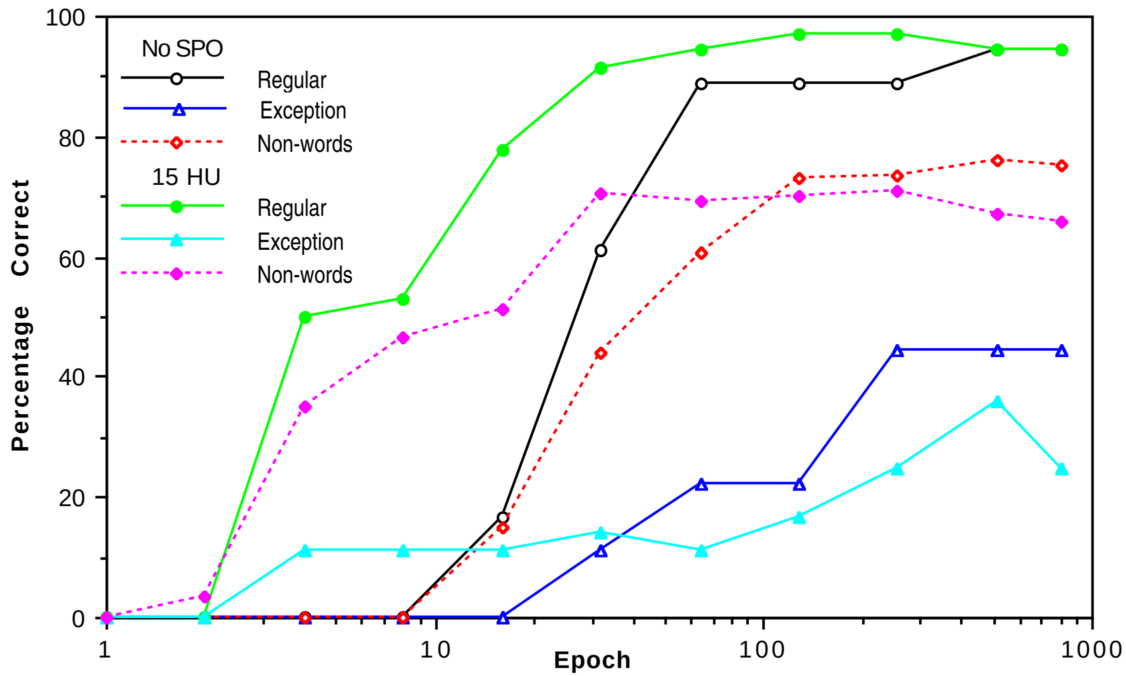
If our neural network models are to provide good accounts of what happens in real brains, we should expect their learning process to be similar to the development in children.



Our networks find regular words (e.g. ‘bat’) easier to learn that exception words (e.g. ‘yacht’) in the same way that children do. It also learns human-like generalization.

Developmental Problems = Restricted Network Learning

Many dyslexic children exhibit a dissociation (performance difference) between regular and irregular word reading. There are many ways this can arise in network models:

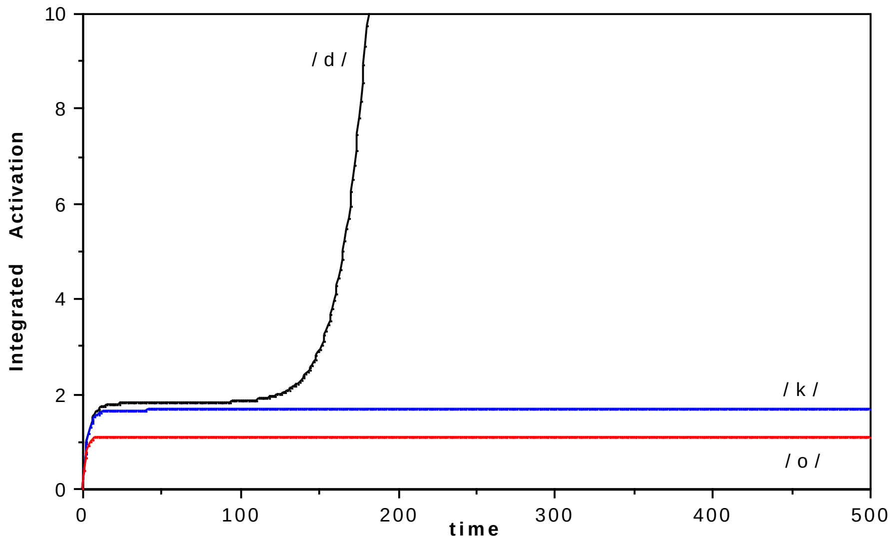


1. Limitations on computational resources (e.g. only 15 hidden units)
2. Problems with learning algorithms (e.g. no SPO in learning algorithm)
3. Simple delay in learning (e.g. low learning rate)

Modelling Reaction Times

Cascaded activation builds up in our output neurons at rates dependent on the network's connection weights. We can thus compute reaction times from our network models:

Reaction Time = Time at Output Action – Time at Input Presentation



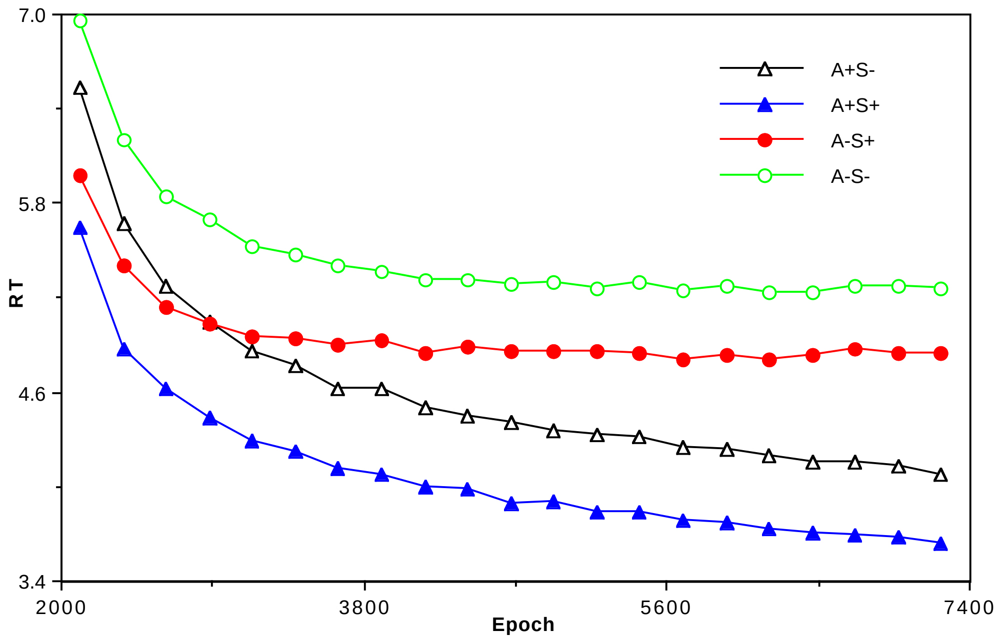
If we present the word 'dog' at the input of our network we can simulate the build-up of output activation for each output phoneme. From these we can determine simulated reaction times for whole words. Generally we average the results over matched groups of words.

High frequency words are pronounced faster than low frequency words. Regular words are pronounced faster than irregular words when they are low frequency, but not when they are high frequency. This is exactly the same pattern found with human subjects!

Modelling Lexical Decision Reaction Time Priming

Semantic priming Semantically related words facilitate lexical decision, e.g. ‘boat’ primes ‘ship’. This arises naturally in neural nets due to over-lapping semantic representations.

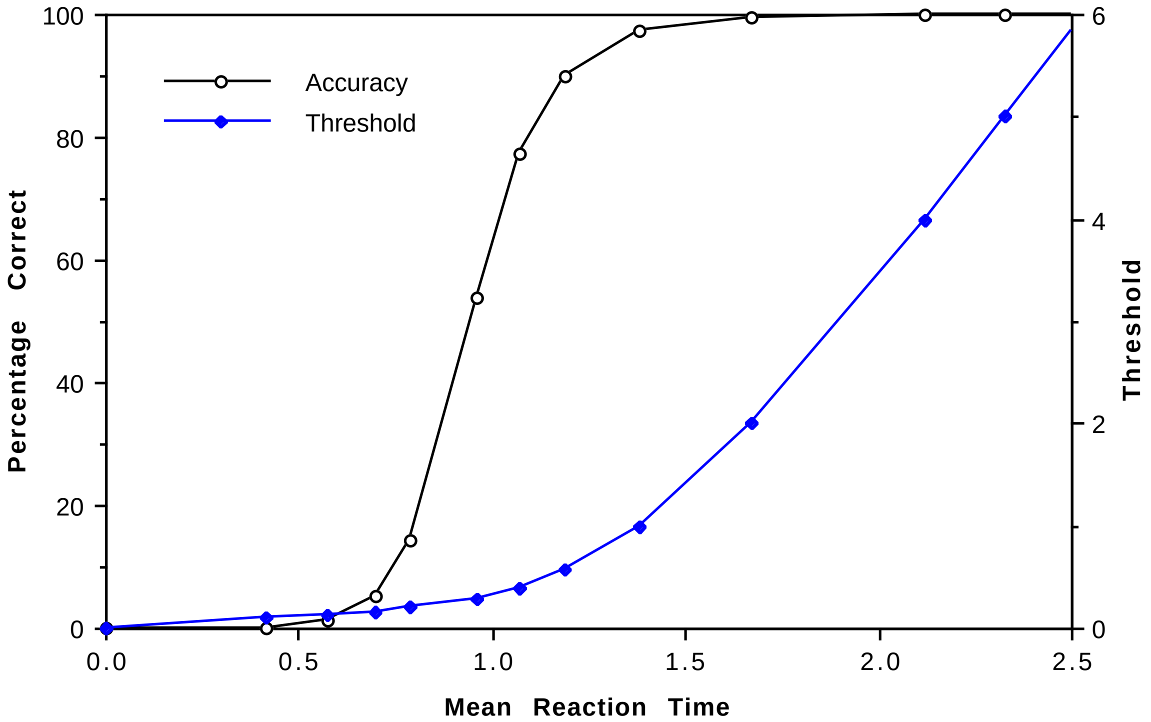
Associative priming Semantically unrelated words can also provide facilitation, e.g. ‘pillar’ primes ‘society’. This will also arise naturally in network models if they can learn that being prepared for common word co-occurrences speeds their average response times.



We can plot the reaction times for a set of words for each different prime (i.e. preceding word) type during training. We can also study the effect of prime duration and target degradation. The pattern of priming results is again in line with that found in human subjects.

Modelling Speed-Accuracy Trade-offs

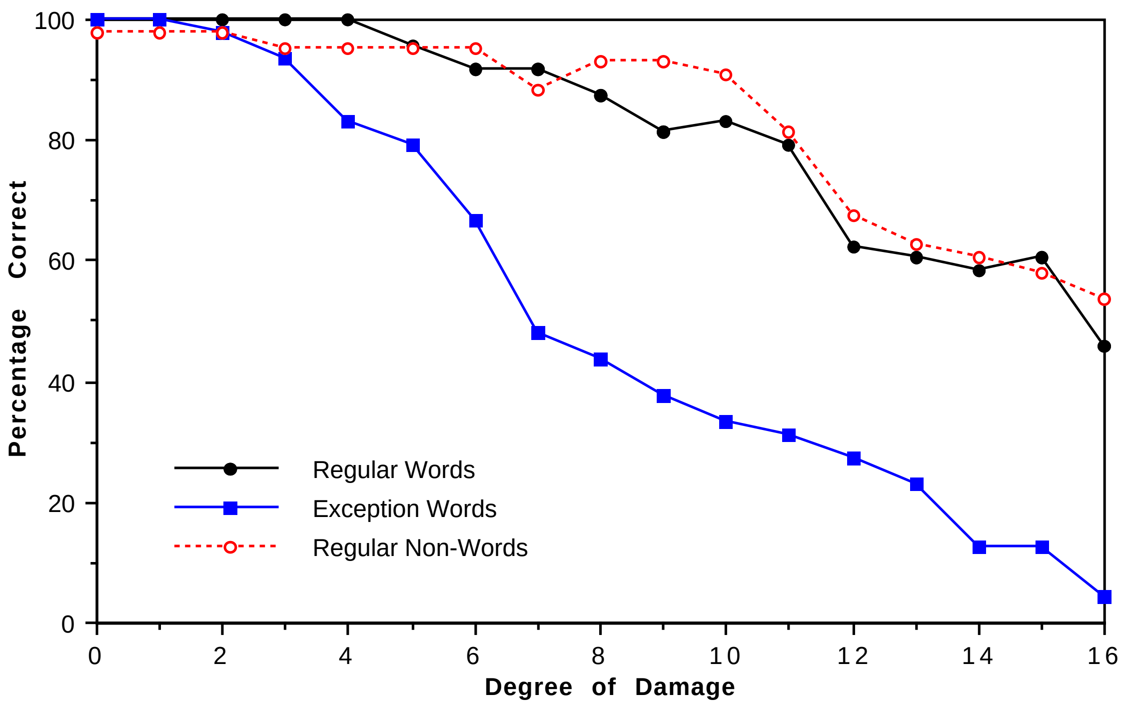
We simulate reaction times by measuring how long it takes for cascaded output activations to build up to particular thresholds in our models. By lowering the thresholds we can speed up the responses, but risk getting the wrong responses. For the reading model:



The sigmoidal shape of the speed-accuracy trade-off curve is very human-like.

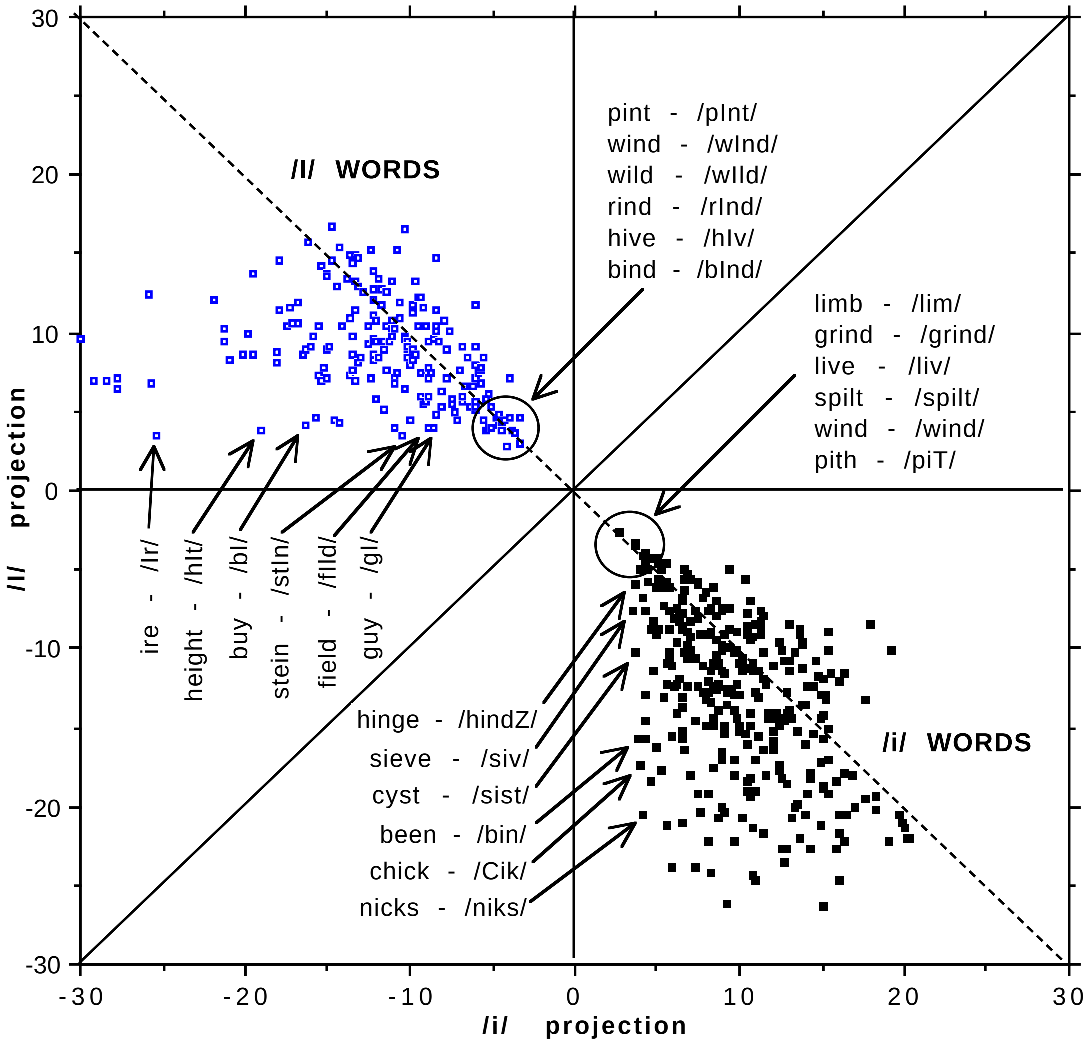
Brain Damage = Network Damage

One advantage of neural network modelling is having natural analogues of brain damage – the removal of sub-sets of neurons and connections, or adding noise to connection weights. If we damage our reading model, the regular items are more robust than the irregulars:



The neural network follows the same pattern as found in human acquired *Surface Dyslexia*.

Internal Representations & Surface Dyslexia



One can look at the representations that the neural network learns to set up on its hidden units. Here we see the weight sub-space corresponding to the distinction between long and short 'i' sounds, i.e. the 'i' in 'pint' versus the 'i' in 'pink'. The irregular words are closest to the border line. So, after net damage, it is these that cross the border line and produce errors first. Moreover, the errors will mostly be regularisations. This is exactly the same as is found with human surface dyslexics.

Modelling More Complex Human Abilities

We have seen how some very simple neural networks can account for a wide range of empirical human data on reading aloud and lexical decision. Neural networks are usually good when fairly simple input-output mappings or control systems are required.

Problems requiring complex reasoning, sequential thought processes, variable binding, and so on, generally prove difficult for neural networks to learn well. Much research is still going on to show how neural networks can, in principle, do such things.

Moreover, it is sometimes just as difficult to understand how our neural networks have learnt to operate, as it is to understand the brain system it meant to be modelling.

In practice, it is often easier to abstract out the essential ideas of the problem, and use a non-neural network (e.g. a symbol processing) approach. This is true both for brain modelling and for artificial system building.

Most of the rest of this module will be concerned with non-neural network approaches to AI. A whole module will be dedicated to neural networks in the Second Year.

Implications for Building AI Systems

Brains have evolved by natural selection to be very good at what they do, so it makes sense to employ the results of that evolution when building AI systems.

The idea is that, since brains exhibit intelligent behaviour, models of brains should also show intelligent behaviour, and consequently be a good source of ideas for AI systems.

Real brains, however, are enormously complex, and our brain models currently capture very little of that complexity. We need to abstract out the essential processes at various levels of description, and work with those.

We must remember, however, that the evolutionary process has itself placed constraints on what can emerge. Birds, for example, must be composed of biological matter, and so feathers are a good solution to the requirements of flying. Aeroplanes made out of metal actually perform much better, and work on very different principles to birds.

While we should clearly make the most of ideas from brain modelling, we should not allow it to restrict what kinds of AI systems we build.

References / Advanced Reading List

1. Bullinaria, J.A. (1997). Modelling Reading, Spelling and Past Tense Learning with Artificial Neural Networks. *Brain and Language*, **59**, 236-266.
2. Bullinaria, J.A. (1997). Modelling the Acquisition of Reading Skills. In: A. Sorace, C. Heycock & R. Shillcock (Eds), *Proceedings of the GALA '97 Conference on Language Acquisition*, 316-321. Edinburgh: HCRC.
3. Bullinaria, J.A. (2002). Lesioned Networks as Models of Neuropsychological Deficits. In: M.A. Arbib (Ed.) *The Handbook of Brain Theory and Neural Networks, Second Edition*, *635-638*. Cambridge, MA: MIT Press.
4. Coltheart, M., Curtis, B., Atkins, P. & Haller, M. (1993). Models of Reading Aloud: Dual-Route and Parallel-Distributed-Processing Approaches. *Psychological Review*, **100**, 589-608.
5. Plaut, D.C., McClelland, J.L., Seidenberg, M.S. & Patterson, K.E. (1996). Understanding Normal and Impaired Word Reading: Computational Principles in Quasi-Regular Domains. *Psychological Review*, **103**, 56-115.
6. Plaut, D.C. & Shallice, T. (1993). Deep Dyslexia: A Case Study of Connectionist Neuropsychology. *Cognitive Neuropsychology*, **10,** 377-500.
7. Shallice, T. (1988). *From Neuropsychology to Mental Structure*. Cambridge, UK: CUP.

Overview and Reading

1. We began by looking at three broad categories of constraints on our brain models – development, adult performance, and neuropsychological deficits.
2. We then saw how some very simple neural network models could account for a broad range of empirical data on reading aloud and lexical decision. The same approaches could be applied to a large range of other human tasks.
3. We ended by looking at the implications this has on building AI systems in general – for more complex brain processes and for real world applications.

Reading

1. The Computational Brain, P.S. Churchland & T.J. Sejnowski, MIT Press, 1994. This is a whole book on computational brain modelling with numerous interesting examples.
2. The items in the Advanced Reading List above all provide examples of brain modelling that may clarify the issues covered in today's lecture.