

# Chapter 1: Approaches to human cognition

## What's it about?

There is more interest than ever in understanding the workings of the human brain and mind. This chapter introduces four main approaches to studying human cognition – cognitive psychology, cognitive neuroscience, cognitive neuropsychology and computational cognitive science. The techniques, strengths and limitations of each approach are discussed.

## What's in this chapter?

- **Cognitive psychology**

**WEBLINK:** [Cognitive science](#)

Historically, most cognitive scientists have adopted the information-processing approach, during which an external stimulus causes internal cognitive processes to occur. These processes produce the desired response:

- Processing directly affected by the input is known as bottom-up processing.
- Serial processing is when only one process occurs at a given time.
- Top-down processing occurs when processing is influenced by the individual's knowledge and expectations.

Most human cognition involves a mixture of bottom-up and top-down processing. Processing in which some or all of the processes involved in a task occur at the same time is known as parallel processing. Cognitive psychologists typically design tightly controlled experiments, which are run on healthy individuals under laboratory conditions. An issue in experimental design is the task impurity problem. This is when a task may rely on a complex mixture of cognitive processes, making interpretation of results difficult. The findings of cognitive psychology have had a major influence on research done in cognitive neuroscience, cognitive neuropsychology and computational cognitive science.

### *Limitations*

People's behaviour in a laboratory may differ from everyday life (i.e., lacking ecological validity). Theories expressed in verbal terms tend to be vague and it is hard to know what predictions follow from them. Findings obtained in a given task are sometimes specific to that task and do not generalise to other tasks – this is paradigm specificity. Emphasis has been on specific theories applicable to a narrow range of cognitive tasks. A comprehensive theoretical architecture linking different components of the cognitive system has been lacking. Cognitive psychology has been of great importance historically. Well-controlled laboratory studies have revealed much about the processes and structures involved in human cognition. However, such studies sometimes lack ecological validity, and may suffer from paradigm specificity. Furthermore, theories in cognitive psychology may apply only to a narrow range of cognitive tasks, and tend to be expressed in rather vague terms.

- **Cognitive neuropsychology**

**WEBLINK:** [Cognitive neuropsychology](#)

Cognitive neuropsychology is concerned with the patterns of cognitive performance shown by patients with brain damage. These patients have suffered lesions – structural alterations in the brain caused by injury or disease. The study of brain-damaged patients can tell us much about normal human cognition.

An example of this approach is patient KF (Shallice & Warrington, 1970), who suffered brain damage and appeared not to have a short-term memory but retained his long-term memory. Coltheart (2010, p. 3) stated that “The principal aim of cognitive neuropsychology is not to learn about the brain. Its principal aim is instead to learn about the mind, that is, to elucidate the functional architecture of cognition.” One key assumption is functional modularity, which implies that cognitive systems consist of sets of modules operating independently of each other:

- Modules exhibit domain specificity. That is, they respond only to one class of stimuli.
- Fodor (1983) argues that humans possess various input modules for encoding and recognising perceptual input, but that the central system (for higher-level processes) is not modular.
- However, others have suggested that most information-processing systems are modular – the “massive modularity hypothesis”.

Another major assumption is that of anatomical modularity – each module is located in a specific and potentially identifiable area of the brain. If modules were distributed across large areas of the brain, then most brain-damaged patients would suffer damage to many modules and there would be few useful data. However, there is little support for the notion of anatomical modularity. Another assumption is known as “uniformity of functional architecture across people”. The assumption of subtractivity suggests that brain damage can impair or delete a function, but cannot introduce new ones.

A crucial goal is the discovery of a dissociation, which occurs when a patient performs normally on one task but is impaired on a second task. A problem with drawing conclusions from this is that it may occur simply because the second task is more difficult. The agreed solution is to find a double dissociation where one patient performs normally on task X and is impaired on Y, and another patient is normal on task Y but impaired on X. Double dissociations are good evidence that two systems are at work. However, there are limitations with the use of double dissociations:

- Patients may have damage to more than one module, and tasks may require more than one module.
- Only some double dissociations have genuine theoretical relevance.
- Double dissociations are not useful for demonstrating three or more separate systems.

An association is when a patient is impaired on task X and also on task Y. Syndromes are certain sets of symptoms usually found together. Unfortunately, associations often tell us nothing about the functional organisation of the brain.

In most psychological research, it is argued that we can have more confidence in our findings when they are based on large groups of participants. However, the group approach is problematic for cognitive neuropsychology because different patients do not show the same impairments. Single-case studies are often used. However, impairments may be due to individual strategy, relative difficulty of tasks or a premorbid gap in function. It may also reflect the reorganised, and not the original, system. Several patients thought to have similar cognitive impairments are tested and then the data of individual patients are compared and variation across patients assessed. This describes a case-series study.

### *Limitations*

Some impact of damage may be camouflaged by compensatory strategies. It would be ideal to find patients in whom brain damage had affected only one module. In practice, however, brain damage is typically more extensive. Much research on cognitive neuropsychology is based on the seriality assumption, which appears to be incorrect. There are often large differences among individuals having broadly similar brain damage. Cognitive neuropsychology has often been applied to relatively specific aspects; for example, language research has included much work on reading and spelling but little on text comprehension.

Cognitive neuropsychologists assume that the cognitive system is modular, that there is isomorphism between the organisation of the physical brain and the mind, and that the study of brain-damaged patients can tell us

much about normal human cognition. Double dissociations are typically much more informative than associations of symptoms and syndromes. It can be hard to interpret the findings from brain-damaged patients for various reasons: they may develop compensatory strategies after brain damage; the brain damage may affect several modules; and there are large differences between groups of individuals with broadly similar damage.

**WEBLINK:** [Michael Gazzaniga in conversation with Shaun Gallagher](#)

- **Cognitive neuroscience: the brain in action**

Cognitive neuroscience involves study of the brain as well as behaviour. Brodmann (1868–1918) produced a cytoarchitectonic map of the brain based on variations in the cellular structure of the brain tissue. Many areas identified by Brodmann have been found to correspond to functionally distinct areas.

**WEBLINK:** [Interactive atlas of the brain](#)

Bullmore and Sporns (2012) argued that two major principles might determine brain organisation:

- the *principle of cost control*:
  - costs would be minimised if the brain consisted of limited, short-distance connections;
- the *principle of efficiency*:
  - efficiency in terms of the ability to integrate information across the brain through numerous long-distance connections.

The brain appears to have a trade-off between these two principles. Van den Heuvel et al. (2009) found IQ did not correlate with the total number of brain network connections. However, there were impressive associations between IQ and the global efficiency of functional brain networks.

## Techniques for studying the brain

Neuroimaging techniques allow us to work out where and when in the brain specific cognitive processes occur. Each technique has strengths and limitations, and techniques vary in their spatial and temporal resolution.

### *Single-unit recording*

Single-unit recording is a fine-grain technique permitting the study of single neurons. A micro-electrode is inserted into the brain to record extracellular potentials. Information about neuronal activity can be obtained over a wide range of time periods, however, it only provides information at the single-neuron level.

### *Event-related potentials (ERPs)*

An electroencephalogram (EEG) is based on recordings of brain activity measured at the scalp, but a key problem is that background activity may obscure the impact of stimulus processing – a solution is to present the stimulus several times to produce event-related potentials (ERPs):

- ERPs have limited spatial resolution but high temporal resolution, providing a continuous measure of the time course of the response.
- ERPs have been used to investigate the time course of processing.
- ERPs are mainly of value when the stimuli are simple and the task involves basic processes occurring at a certain time after stimulus onset.

**WEBLINK:** [An overview of EEG and MEG](#)

### *Positron emission tomography (PET)*

Positron emission tomography (PET) is based on detection of positrons from radioactively labelled water injected into the body. Increased blood flow to active areas of the brain is indicated by increased positron detection:

- PET has reasonable spatial resolution (5–10mm) but very poor temporal resolution (30–60s).
- PET provides only an indirect measure of neural activity and is an invasive technique.

**WEBLINK:** [An overview of PET scans](#)

#### *Magnetic resonance imaging (MRI and fMRI)*

In magnetic resonance imaging (MRI), radio waves are used to excite atoms in the brain, which produces magnetic changes detectable by a large magnet surrounding the body:

- MRI produces a precise 3-D picture.
- MRI only tells us about structure, not function.
- fMRI has a better temporal resolution (2–3s) and spatial resolution (1mm) than PET.

Functional MRI (fMRI) measures the BOLD (blood oxygen level-dependent contrast) signal. This is a distortion in the local magnetic field when oxyhaemoglobin is converted to deoxyhaemoglobin when neurons consume oxygen. Event-related fMRI (efMRI) allows comparison of brain activity across different cognitive events, for example during learning and retrieval in a recognition memory task. fMRI is more useful than PET as it provides more precise spatial information and shows changes over shorter periods of time. However, fMRI provides an indirect measure of neural activity. The experience may be uncomfortable for participants, and there are constraints to the types of tasks that can be used in the scanner.

**WEBLINK:** [How fMRI works](#)

#### *Magneto-encephalography (MEG)*

Magneto-encephalography (MEG) uses a super-conducting quantum interference device (SQUID) to measure magnetic fields produced by electrical brain activity:

- MEG has excellent temporal resolution and very good spatial resolution.

#### *Transcranial magnetic stimulation (TMS)*

In transcranial magnetic stimulation (TMS), a coil is placed close to the participant's head and a brief but large pulse of current is run through it in a short time. In practice several pulses are usually administered (repetitive transcranial magnetic stimulation (rTMS)). Each magnetic pulse creates a short-lived magnetic field that inhibits processing activity in the area:

- TMS produces a “temporary lesion” enabling the role of a brain area in a given task to be assessed.
- TMS has exciting potential and can be used to make causal statements about brain areas necessary for normal task performance.
- TMS has been used to demonstrate the involvement of the dorsolateral prefrontal cortex in task coordination (Johnson & Zatorre, 2006; see Eysenck & Keane, 2010, p. 13), as well as when a brain area is most involved in task performance. An advantage of TMS over neuroimaging is that it allows us to evaluate the necessity of brain areas for task performance. Compared to research with brain-damaged patients, the brain areas involved in TMS “lesions” are tightly controlled, and TMS does not lead to development of compensatory strategies. However, limitations of TMS are that:
  - its effects are complex;
  - it can only be applied to brain areas with no overlying muscle;
  - it may cause activity changes in brain areas distant from the stimulation site;
  - there are possible safety issues (e.g., seizures).

### *Evaluation*

Neuroimaging reveals associations between patterns of brain activation and behaviour rather than direct evidence concerning cognitive processing. Most functional neuroimaging research is based on the assumption of functional specialisation. This assumption may not hold true for higher-order cognitive functions involving greater integration and coordination across the brain. Functional neuroimaging may have direct relevance to resolving theoretical controversies within cognitive psychology. For example, Kosslyn and Thompson (2003) found in a meta-analysis of neuroimaging studies that similar processes are used in visual imagery and perception. Martinez et al. (1999) found in an ERP study that unattended stimuli are less thoroughly processed than attended stimuli. There is the issue of defining an appropriate baseline with which to compare brain activity during task performance. Brain activity may increase or decrease when compared to the baseline. Research in cognitive neuroscience also suffers from a lack of ecological validity, and from paradigm specificity. For example, long-term recognition memory is worse in the fMRI environment (Gutches & Park, 2006).

Cognitive neuroscientists use various techniques for studying the brain, and these techniques vary in their spatial and temporal resolution. It is often useful to combine two techniques that differ in their strengths. Functional imaging reveals associations between patterns of brain activation and behaviour rather than providing direct evidence of cognitive processing, or of causality. Transcranial magnetic stimulation has the advantage that it can potentially be used to show that a given brain area is necessarily involved in a particular cognitive function. Functional imaging data require careful interpretation because functional specialisation is not always found (particularly for higher-order functions) and there are issues with defining an appropriate baseline, as well as ecological validity and paradigm specificity. However, cognitive neuroscience has contributed to the resolution of important theoretical issues, and continues to shed light on the relationship between brain and behaviour.

**WEBLINK:** [A visual overview of imaging techniques](#)

**WEBLINK:** [The homepage of the Cognitive Neuroscience Society](#)

- **Computational cognitive science**

**WEBLINK:** [AI on the web](#)

Computational cognitive science is an approach to understanding human cognition that uses computational models. Computational modelling involves programming computers to mimic aspects of human cognitive functioning. Artificial intelligence, in contrast, involves constructing computer systems that produce intelligent outcomes, but the processes may bear little resemblance to those used by humans.

A good computational model can show us that a given theory can be specified and allows us to predict behaviour in new situations. A major benefit of computational models is that they can provide an explanatory and predictive basis for a phenomenon. Implementing a theory as a program is a good method for checking that it contains no hidden assumptions or vague terms.

Palmer and Kimchi (1986) argued that it should be possible to decompose a theory successively through a number of levels (from descriptive, to a written program). Other issues arise about the relationship between the performance of the program and human performance (Costello & Keane, 2000)

### *Connectionist networks*

**WEBLINK:** [Explanations of connectionism](#)

Connectionist networks are also called neural networks or parallel distributed processing (PDP) models. The network consists of elementary neuron-like units, or nodes, connected together:

- Units affect other units by excitation or inhibition.
- Units take the weighted sum of all inputs and produce an output if a threshold is exceeded.

- The network is characterised by the properties of the units, the way they are connected, and the rules used to change the strength of connections.
- Networks can have different structures or layers.
- A representation of a concept can be stored in a distributed manner by a pattern of activation.
- The network can store many patterns without them necessarily interfering with each other.

An important learning rule is back-propagation of errors, which is a mechanism allowing a network to learn to associate a particular input pattern with a given output by comparing actual responses against correct ones. The model can be made to learn the behaviour, rather than the behaviour being explicitly programmed. These networks can model cognitive performance without the explicit rules of production systems. For example, the network NETalk was able to learn the rules of English pronunciation without having explicit rules for combining and encoding sounds (Sejnowski & Rosenberg, 1987).

Many (but not all) connectionist models are based on the assumption that knowledge is represented in a distributed fashion, but there is also evidence for local representation of knowledge.

Localist connectionist models include the reading model of Coltheart, the TRACE model of word recognition, and models of speech production. It is likely that some knowledge is represented locally and some is distributed.

#### *Production systems*

Newell and Simon (1972) first established the usefulness of production system models in characterising cognitive processes such as problem solving and reasoning:

- Production systems have numerous THEN production rules.
- They have a working memory.
- They operate by matching the contents of working memory against the IF-parts of the rules and executing the THEN-parts.
- There may be a conflict-resolution strategy, selecting one rule as the best to be executed.
- Many aspects of cognition can be specified as sets of IF ... THEN rules.

## **ACT-R**

Anderson's (1993, [2004](#)) Adaptive Control of Thought-Rational model is a cognitive architecture – a cognitive model that is domain-generic and encompasses a wide range of cognitive applicabilities.

ACT-R is based on the assumption that the cognitive system consists of several modules. Each module has an associated buffer, and a central production system detects patterns in these buffers and takes coordinated action.

**WEBLINK:** [ACT-R website](#)

[Anderson and Lebiere \(2003\)](#) evaluated connectionism and production systems. The pattern of strengths and weaknesses differs between connectionist models and ACT-R ratings differed between the two theories on criterion of operating in human time, using language, accounting for developmental phenomena, and mapping on to the brain. The focus in many connectionist models is on language and cognitive development, and there have been systematic attempts to take some account of actual brain functioning. In contrast, ACT-R has emphasised the importance of predicting human processing time as well as human behaviour.

**WEBLINK:** [Comparison of production and connectionist systems \(also some interesting AI articles\)](#)

#### *Evaluation*

##### **Strengths**

- Computational cognitive science requires theorists to think carefully and rigorously.
- The development of cognitive architectures offers the prospect of an overarching framework for cognition.
- Connectionist networks can demonstrate “learning” to some extent.

- Many connectionist models are based on the assumption that knowledge is represented in a distributed fashion.
- The scope of computational cognitive science has expanded to include modelling of functional neuroimaging data.
- Computational cognitive science can provide powerful theoretical accounts of parallel processing systems.

### **Limitations**

- Computational models are rarely used to make new predictions.
- Computational cognitive scientists typically develop one model of a phenomenon rather than exploring many models.
- Connectionist models that claim to have neural plausibility do not really resemble the human brain, for example in terms of the interconnectivity of the cortex.
- Parameters or variables can be adjusted to produce numerous possible outcomes.
- Computational models often fail to capture motivational and emotional factors.

Computational cognitive scientists focus on computational models, in which theoretical assumptions have to be made explicit. These models are expressed in computer programs, which should produce the same outputs as people when given the same inputs. Production systems and connectionist networks are important types of computational models having somewhat different patterns of strengths and limitations. Production systems consist of productions in the form of “IF ... THEN” rules. Connectionist networks differ from previous approaches in that they can “learn” from experience (e.g., through the back-propagation of errors) and they lack explicit rules. Such networks often have several structures or layers (e.g., input units, intermediate or hidden units, and output units). Representations in connectionist networks are often distributed, but there are also localist connectionist networks. Computational models rarely make new predictions, they lack neural plausibility in several ways, and motivational and emotional influences on cognitive processing are ignored.

### • **Comparison of major approaches**

An increasing amount of research involves two or more approaches. For example, Rees et al. (2000) used both cognitive neuropsychology and functional neuroimaging to study the disorder of extinction and found that extinguished stimuli still produce activation within the visual cortex. Each approach makes its own distinctive contribution, so all are needed and each approach has its strengths and limitations. It is optimal to make use of converging operations in which several different research methods are used to address a given theoretical issue. The strength of one method balances out limitations in the other, producing stronger evidence.

Each approach to studying cognition has its set of strengths and limitations. However, each approach makes a distinct contribution and all are needed. By combining two or more approaches to address the same theoretical issue, stronger evidence can be produced.

### **Additional references**

- Anderson, J.R. (1993). *Rules of the mind*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Anderson, J.R. & Lebiere, C. (2003). The Newell Test for a theory of cognition. *Behavioral and Brain Sciences*, 26: 587–640.
- Fodor, J.A. (1983). *The modularity of mind*. Cambridge, MA: MIT Press.
- Palmer, S.E. & Kimchi, R. (1986). The information processing approach to cognition. In T. Knapp & L.C. Robertson (eds), *Approaches to cognition: Contrasts and controversies*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Rees, G., Wojciulik, E., Clarke, K., Husain, M., Frith, C. & Driver, J. (2000). Unconscious activation of visual cortex in the damaged right hemisphere of a parietal patient with extinction. *Brain*, 123: 82–92.
- Sejnowski, T.J. & Rosenberg, C.R. (1987). Parallel networks that learn to pronounce English text. *Complex Systems*, 1: 145–68.