Memory
From A to Z

Keywords, Concepts, and Beyond

Yadin Dudai
The Weizmann Institute of Science, Rehovot, Israel

OXFORD
UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS

Great Clarendon Street, Oxford OX2 6DP

Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship,
and education by publishing worldwide in

Oxford New York
Auckland Bangkok Buenos Aires Cape Town Chennai
Dar es Salaam Delhi Hong Kong Istanbul Karachi Kolkata
Kuala Lumpur Madrid Melbourne Mexico City Mumbai Nairobi
São Paulo Shanghai Taipei Tokyo Toronto

and an associated company in
Berlin

Oxford is a registered trade mark of Oxford University Press
in the UK and in certain other countries

Published in the United States
by Oxford University Press Inc., New York

First published 2002

British Library Cataloguing in Publication Data

Data available

Library of Congress Cataloging in Publication Data

ISBN 0 19 850267 2

10 9 8 7 6 5 4 3 2 1

Typeset by Cepha Imaging Pvt. Ltd., Bangalore, India

Printed in Great Britain

on acid-free paper by T.J. International, Padstow, Cornwall, UK

Contents

The conceptual framework

The premises that underlie the selection of entries, the adaptation and formulation of definitions, and the views expressed in this book.

I am a functionalist[1] with a biologist's *bias and with *conscious awareness of other disciplines. My approach to memory research is guided by the following tenets: (a) the function of the brain is to create and retain *internal representations of the world that could guide behaviour; (b) the function of *learning is to permit the adaptation of internal representations to a changing world (*memory is the retention of these adaptations over time); (c) learning and memory require neural *plasticity for their actualization; and (d) learning and memory are *system properties, made possible by the concerted operation of multiple *levels of the system.

The aforementioned tenets yield two important consequences for memory research. First, the comprehensive investigation of the processes and mechanisms of biological learning and memory requires a multilevel approach. Second, in the analysis of learning and memory, two levels of functional organization are particularly critical. One is the behavioural level. It does not make sense to address the function of the system without addressing its input–output relationships. The other is the level in which the specific content (semantics) of internal representations emerges in the brain. Identification of the behavioural level is self-evident. Identification of the level that encodes internal representations is not. It is currently believed that the level critical for encoding the semantics of internal representations in the brain is the circuit level, or the cellular-and-circuit level. More reduced levels implement plasticity, but in the absence of the circuit *context, do not suffice to endow the representation with its semantics. It is essential, therefore, that research programmes on memory never lose sight of the circuit and the behavioural levels. This is not easy. The circuit level is often excessively complex, the behavioural level amazingly tricky. Furthermore, the remarkable success of molecular neurobiology is enticing. I thus believe in a focused, restrained *reductionistic approach to memory research (Dudai 1992). I hope that this is aptly reflected in the entries throughout this book.

Each entry opens with a definition, or a set of definitions. What a definition is, is extremely difficult to define. A liberal list contains no less than 18 different species of definitions, and multiple candidate definitions of the definitions in each species (Robinson 1954). Whenever possible, I tried to adhere to one of the following meanings of definition: (a) the minimal set of attributes that uniquely describes an item or a concept; and (b) the formulation of a thing in terms of a more elementary level of organization or theory. These meanings are not mutually exclusive, and reflect, respectively, an attempt to adhere to *Ockham's razor, and the basic reductionistic approach, which has been restrained above. It is evident, however, at the outset, that each of these types of definitions requires quite a lot of *a priori knowledge about the item to be defined.[2] In the case of many items and concepts in the field of memory research, the relevant knowledge is yet unavailable. I had, therefore, to use an additional type of definition: explanation of the meaning of the term as it is to be used (*stipulated definition*). And as terms in memory research are occasionally used in more than one way, I provided multiple definitions when appropriate. The difficulties and uncertainties involved in definitions bring to mind the view that attempts to define entities at the cutting edge of knowledge could cause more harm than good: 'For when we define, we seem in danger of circumscribing nature within the bounds of our own notions' (Burke 1757). There is, however, the opposite view, that the risk is well worth taking. Socrates leads Meno to admit that definitions are always a must for a fertile, constructive dialogue (Plato, *Meno* 79d; *culture). In this debate, while being aware of Burke's caveat, I am much in favour of Meno's conviction.

Each entry ends with a short string of *associations. Bodies of knowledge in general are associative systems. I tried to *reinforce this notion by proposing selected associations. The reader is invited to form additional ones. Associations are not only aids to understanding, they are also proven *mnemonic devices: the richer the associative network, the higher the probability that the item will be stored (*metaphor) and *retrieved.

[1]Functionalism in its broadest sense is any view that analyses something in terms of how it functions (Lacey 1996). There are several versions of functionalism, one of which is 'functional analysis' (Cummins 1975). This is the research strategy that relies on the decomposition of a *system into its component parts while attempting to explain the working of the system in terms of the capacities of the parts and the way they are integrated with each other (Block 1980). Still, the structure of the parts and of the integrative system matters solely as much as it implements or shapes the function. Functional analysis is the sense of functionalism implied here.

[2]On this difficulty, which is also called 'the problem of the criterion', see *criterion.

A Priori

1. Independent of experience.

2. Beforehand.

A priori it could be assumed that students and aficionados of memory will benefit from contemplating the concept of 'a priori'. Before defending the aforesaid statement, however, a brief clarification of the different meanings and uses of 'a priori' is appropriate.

Prior to the eighteenth century, the pair of terms 'a priori'/'a posteriori' (Latin for 'from what is earlier'/'for what comes after') was used to distinguish between modes of reasoning: 'The mind can discover and understand the truth … by demonstration. When the mind reasons from causes to effects, the demonstration is called a priori; when from effects to causes, the demonstration is called a posteriori' (Arnauld 1662). Only later were these nonidentical terminological-twins used to refer to types of knowledge: knowledge independent of experience is 'a priori', that which is grounded in experience is 'a posteriori' (Kant 1781). Traditionally since then, the pair 'a priori'/'a posteriori' is associated in the philosophical discourse with two other pairs of opposites: 'analytic' vs. 'synthetic', and 'necessary' vs. 'contingent' (Moser 1987; Grayling 1997). A statement is 'analytic' if its truth value can be determined by understanding the concepts or terms contained in it, whereas it is 'synthetic' if in order to determine its truth value we must know how the concepts or terms involved relate to other constituents of the world. Hence, adapting a commonly used illustration, 'singles are unmarried' is analytic, because 'single' is 'unmarried', whereas 'singles are happy' is synthetic, because it is not evident from 'singles' how their mood should be (the latter statement also demonstrates that some kinds of truth are *context specific or in the eye of the beholder, but this is another story). In formal terms, an analytic statement is thus a tautology, and its truth value follows necessarily. The latter property leads us to the third related pair of opposites: 'necessary' vs. 'contingent'. 'Necessary' refers to statements that must be either true or false due to what they state, whereas in 'contingent' statements the truth value is contingent upon other occurrences or relationships in the world. Discussion of the 'necessary'/'contingent' pair is within the realm of metaphysics, the 'analytic'/'synthetic' pair deprives logicians of sound sleep, whereas 'a priori'/'a posteriori' is within the domain of epistemology (the science of knowledge) (Moser 1987; Grayling 1997; Bealer 1999).

It is the epistemological connotation of 'a priori' that interests us here. Furthermore, we focus on only a limited portion of the universe: the *individual* organism, its brain, behaviour, and memory. Construing 'experience' in definition 1 as any behavioural or physiological experience of the individual, leaves only one source of a priori knowledge in the individual brain: the genetic material. Genes carry information about a variety of behavioural capabilities and capacities (*neurogenetics). This information is hence 'innate'.[1] As far as the individual is concerned, this is *bona fide* a-priori knowledge. For the species it is not, because the knowledge is supposed to have been acquired over time, a posteriori, by natural selection in evolution. However, it is also useful to consider as 'a priori' that knowledge that cannot be explained *solely* by the individual's experience. Such knowledge is generated by *developmental processes, via the interaction of genes and environment in prenatal and early postnatal periods. It is also produced throughout life by the endogenous activity of the brain, which depends on the processing of both innate and acquired knowledge. Definition 2 is colloquial: according to it, 'experience' is 'experience at the present time', e.g. while on a learning task. Hence according to this liberal interpretation any experience provides a priori knowledge for future experiences. This connotation of a priori gravitates toward the trivial, and will not be further discussed here.

A priori knowledge of both innate and postnatal origin fulfils multiple roles in behaviour and behavioural *plasticity:

1. Innate knowledge underlies reflexes and predetermined behavioural routines such as used in feeding, mating, fighting, and fleeing (Lorenz 1981; Dudai 1989). These behaviours vary in their dependency on postnatal experience. Some are essentially independent of experience, although they still may be perfected or modified by it, e.g. α-type *classical conditioning. Other behaviours require experience for maturation, fine tuning, and optimal *performance. This experience may have to be provided during a restricted 'sensitive period' in life, as in *imprinting (Lorenz 1981) and *birdsong (Nelson and Marler 1994). Another, more general type of 'prepared' or 'constrained learning', in which the type of associations, but not their actual content, is constrained a priori, is *conditioned taste aversion: we are inclined a priori to associate the taste of foodstuff with subsequent visceral malaise but not with a painful blow to the skin (Garcia et al. 1968). Admittedly, most philosophers would not like the use of the term 'knowledge' in the context of such 'simple' behaviours: 'No philosopher will be disturbed if Lorenz

tells him that young geese follow the farmer around without previous conditioning or training. If Lorenz were to add that the young goose knows that it should follow the farmer, or that the farmer is a friend, philosophical ears would be pricked' (Cooper 1972). However, first of all, 'knowledge' is here used in its most *reductive connotation, not necessarily involving *conscious awareness (*internal representation); second, irrespective of the status of philosophical ears, the question whether animals are 'consciously aware' or not is not yet settled (*declarative memory).

2. Innate knowledge underlies capacities and operational rules of higher brain faculties such as language and mathematical abstraction in humans ('the speaker of a language knows a great deal that he has not learned', Chomsky (1966); compare Socrates on geometry: 'Try to discover by recollection what you do not know, or rather what you do not remember', Plato, *Meno* 86b).

3. Perhaps most intriguing is the notion that a priori knowledge that draws from a combination of innate and acquired resources permits our brain to anticipate the world on a momentary basis (e.g. Anokhin 1974). This issue relates to one of the most profound problems in the neurosciences and the philosophy of mind: the relationships of internal representations to the outside world. Let us consider two basic possibilities. One is that input from the world somehow *instructs* the brain to generate specific internal representations of reality. This type of process does not necessitate a priori knowledge, although it may still benefit from it. The other possibility is that the world somehow *selects* representations among 'pre-representations', which are generated endogenously in the brain (Young 1979; Heidmann *et al.* 1984; Dudai 1989; Edelman 1993; *stimulus). The 'selectionist' view has a Darwinian flavour, and likens the ontogenesis of our mind to the phylogenesis of our species. According to this view, the mammalian brain is not a passive observer but rather an active agent that anticipates the immediate future (*planning), and toward that end keeps itself busy by generating internal *models of reality. The postulated rules that guide 'the survival of the fittest internal models' may take into account predictions based on both innate knowledge and accumulated experience, and congruency with the on-line demands of the real world as conveyed by the senses. Such capacity is hence expected to be subserved in every individual of the species by two tiers of a priori knowledge. First there are the species-specific innate components responsible for much of the rules and the hardware, namely the computations, *algorithms and neuronal devices that enable the brain to generate and stabilize the aforementioned pre-representations (*level). Then there is the ongoing flux of the short-lived pre-representations themselves, which are unique to each individual of the species, and could be regarded as flashes of subjective knowledge preceding *perception and the *acquisition of memories. In this case, the past literally chases the present, and 'a priori' may refer to a time-scale of seconds only. Still, this is 'a priori', because at least part of the information is not derived from actual experience in the outside world.

The 'selectionist' hypothesis hence implies that we continuously anticipate the world and generate approximate models of it, and that both endogenous and exogenous information combine to represent reality (e.g. Arieli *et al.* 1996). This raises the question how faithful to reality are our internal representations (*false memory, *real-life memory). We may assume that in the course of evolution, our ability to model the world, learn about it, and interact with it has been shaped to reach a reasonable correspondence of the internal models to reality. The fact that organisms succeed in negotiating with an ever changing milieu attests to that. But not all our memory *systems (*taxonomy) have been subjected to the same selective pressures, such as the pressure for improved precision and detail. Hence, whether a specific type of memory, such as *declarative, is inherently faithful to reality or not, is itself a priori influenced by evolutionary forces.

Last, we should not *forget that in daily life we are all constrained by a priori assumptions that could *bias our personal (or *cultural) attitude toward events, facts, and disciplines. The attitude toward 'memory' is not expected to be an exception.

Selected associations: Acquisition, Bias, Development, Palimpsest

[1]For *classic philosophical attitudes to innate knowledge in general, see Locke (1690) and Leibniz (1704).

Acetylcholine

A *neurotransmitter at central *synapses and at the vertebrate neuromuscular junction.

Acetylcholine (ACh; the acetic acid ester of choline) was among the first chemicals to be proposed as a neurotransmitter, and the first neurotransmitter to be identified in and isolated from neural tissue (Dale 1914; Loewi 1921). It was also the first for which the

existence of a proteineous membrane *receptor had been suggested (Nachmansohn 1959).[1] *In vivo* ACh is synthesized from the amino alcohol choline and acetyl coenzyme A. The job is done by the enzyme choline acetyltransferase (Kitamoto *et al.* 1992). ACh is hydrolysed by another enzyme, acetylcholinesterase, one of the fastest enzymes ever (Taylor and Radic 1994). Receptors for ACh are of two major types:

1. 'Nicotinic', so-called because they bind nicotine (the tobacco poison). Nicotinic receptors are *ion-channel receptors, i.e. they contain a pore that mediates the flux of ions across the membrane and is gated by the neurotransmitter (Karlin and Akabas 1996).
2. 'Muscarinic', so-called because they bind muscarine (a mushroom poison that kills flies, *Musca*). Muscarinic receptors are 'metabotropic', i.e. they do not include a channel but rather exert their effect by modulation of *intracellular signal transduction cascades (Wess 1993).

Each of these receptor types can be further classified into subtypes. The subtypes are commonly characterized by their affinity and specificity for activators (agonists) and inhibitors (antagonists); the identity of the intracellular signal transduction cascades coupled to the receptor; and the cellular localization (presynaptic or postsynaptic).

A neuronal *system in which ACh is a neurotransmitter or neuromodulator is termed 'cholinergic'. Cholinergic innervation of various brain areas such as the *cerebral cortex could be described as either extrinsic, e.g. stemming from central cholinergic nuclei in the brain, or intrinsic (Johnston *et al.* 1981; Mesulam *et al.* 1983). The central cholinergic nuclei in the mammalian brain are located in the basal forebrain and the brainstem (Figure 1). The major ones are in the basal forebrain and they innervate the neocortex, *hippocampus, and parts of the *amygdaloid complex. Those in the brainstem innervate among other the thalamus. The innervation by the central cholinergic nuclei is an example of a 'diffused neuromodulatory system' i.e. a neuromodulatory system that does not target specific synapses or neurons but rather a whole region or multiple regions (see also *dopamine, *noradrenaline).

The cholinergic basal forebrain system, itself a collection of nuclei, has been repeatedly implicated in cognition, including *attention, learning, and memory. A correlation was found in a number of studies between degeneration of basal forebrain nuclei, cholinergic dysfunction and cognitive deterioration in Alzheimer's disease (*dementia) and in aged humans and rodents. This has led to the 'cholinergic hypothesis of memory dysfunction' (Bartus *et al.* 1982). This hypothesis proposes that cholinergic dysfunction is not only a correlate, but also a cause of cognitive and behavioural deficits in dementia. The 'cholinergic hypothesis' was highly successful at least on one front: it generated a surge of research on the potential role of cholinergic modulation in learning and memory, and served as an incentive for the development of cholinergic drugs to treat dementia (see below).

Multiple processes and mechanisms have been suggested to underlie the postulated roles of ACh in

Fig. 1 A schematic diagram of the central cholinergic projections in the mammalian brain. There are two major projectional networks: from the basal forebrain, innervating among others the *cerebral cortex (CTX), *hippocampus (HIP) and *amygdala (AM); and from the penducolopontine and laterodorsal tegmental nuclei (marked in the figure as PPT), innervating among others the thalamus (TH) and tectum (TEC). OB, olfactory bulb. Local cholinergic circuits are not shown. (Adapted from Cooper *et al.* 1996.)

learning and memory. As is the case with other neurotransmitters and neuromodulators, the physiological roles of ACh in brain should be judged not only by its independent activation of specific cellular receptors and their downstream intracellular signal transduction cascades, but also by its contribution to the activation and cross-talk of webs of signalling cascades induced by coactive sets of neurotransmitters and neuromodulators (*coincidence detector, *context). Similarly, at the circuit *level, the function of the cholinergic system must be assessed in the context of the concerted activity of multiple neurotransmission and neuromodulatory pathways on the target circuit (Decker and McGaugh 1991). ACh was portrayed as a cellular code for saliency (*surprise), *attention, *state dependency, and even as a direct 'storage signal' that instructs the appropriate circuits to encode novel information as lasting *internal representations (Mishkin and Murray 1994; Naor and Dudai 1996; Everitt and Robbins 1997; Wenk 1997; Shulz *et al.* 2000). All the above functions could actually be different manifestations of similar cellular and circuit mechanisms, with the specific role of the cholinergic function in a given cognitive and behavioural situation being dependent upon the task, the context, and the identity of the brain areas involved. At the *algorithmic level, brain ACh, similarly to other neuromodulators such as *noradrenaline, was proposed to enhance the signal-to-noise ratio in the target circuit (Barkai and Hasselmo 1997).

In recent years, the function of ACh in the mammalian brain has been scrutinized by a variety of novel *methodologies, techniques, and preparations. Not all the data so obtained fit smoothly into the hypothesis that ACh is indeed obligatory for learning, certainly not in all types of learning, but the overall picture favours the idea that it does play an important part in many learning situations. A somewhat surprising finding was reported by several laboratories following the introduction of a powerful experimental tool, the chimera-immunotoxin 192IgG-saporin. This toxin is a synthetic chimera between the toxin saporin, that kills cells, and an antibody to a subtype of a receptor for nerve growth factor that resides on most types of cholinergic neurons in the basal forebrain. The compound guides itself to these cholinergic neurons and destroys them selectively, while leaving other neurons, the majority of which are noncholinergic, intact. In disparity with the effect of less selective lesions of basal forebrain cholinergic nuclei, in several preparations, the guided toxin had only a small effect if at all on memory (e.g. Baxter *et al.* 1995; but see, for example, Power *et al.* 2002). In contrast, a variety of other new experimental manipulations did support a correlative and in certain cases an obligatory role of ACh, acting either via muscarinic or via nicotinic receptors, in a variety of learning situations and of neuronal *plasticity mechanisms that *model attention and learning (Auerbach and Segal 1996; Gray *et al.* 1996; Picciotto *et al.* 1998; Berman *et al.* 2000; Mansvelder and McGehee 2000; Nail-Boucherie *et al.* 2000; Rasmusson 2000; Shulz *et al.* 2000). For example, in many preparations, ACh enhances transmitter release, and in some it supports *long-term potentiation. Stimulation of the basal forebrain cholinergic input was shown to enable the reorganization (*plasticity) of cortical sensory *maps, and hence possibly *internal representations, in response to modality-specific input (Bjordahl *et al.* 1998; Kilgard and Merzenich 1998); a caveat is, however, appropriate regarding such an approach, because, as noted above, the basal forebrain is also a source of noncholinergic innervation to the cortex. Another report that made it to the headlines was that transplantation into the brain of cells engineered to release ACh alleviates cognitive deficits in rats with a cholinergically denervated cortex (Winkler *et al.* 1995).

A good deal of support for the role of the cholinergic system in cognition stems from human pharmacology. Drugs that increase the availability of ACh, mostly inhibitors of acetylcholinesterase, have beneficial effects on cognitive function at the early stages of dementia. Furthermore, to the understandable dismay of nonsmokers, nicotine appears to be moderately beneficial to attention and memory (Di Carlo *et al.* 2000). It thus appears that cholinergic drugs establish themselves as cognitive boosters (*nootropics) before the exact and task-specific roles of ACh in cognition and memory are fully understood. This, of course, is not unique to the cholinergic drugs; if understanding the mechanism of action was a *criterion for the introduction of a drug, many of our most efficient medications would not be in use.

Selected associations: Attention, Dementia, Neurotransmitter, Receptor, Synapse

[1]For an early suggestion that there should be a receptor, long before ACh itself was discovered, see Langley (1878).

Acquisition

1. The initial *phase in the formation of a *memory trace.

2. The process by which new information is converted into a memory trace.

3. The change in *performance during training that is taken to represent the progression of *learning.

Memories are like people—they are born, live, and die. Acquisition is their moment of birth. The other major phases in the life history of a memory are *consolidation (if it is ever to become a long-term memory), storage, *retrieval, and extinction (*experimental extinction, *forgetting). Depending on the context of discussion, 'acquisition' implies a temporal phase (definition 1, e.g. Stillings *et al.* 1987); or a process that takes place during this phase (definition 2, e.g. Tulving 1983); or a change in *performance that reflects this process (definition 3, e.g. *behaviourism). This change in performance is quantified by an 'acquisition curve' or 'learning curve', in which performance is plotted against the amount of practice (e.g. Skinner 1938; e.g. Figure 41, p. 144). Commonly, the *subject is said to have completed the acquisition of the task if its performance has reached a preset *criterion, such as time to reach the goal in a *maze or a certain probability of success on a discrimination problem (e.g. *delay task). The process of acquisition was termed 'engraphy' by Semon (1904), meaning the engraving of an *engram, but 'engraphy' has never caught on. 'Acquisition' is sometimes used as a synonym for '*learning', but the latter term has a broader meaning and usage.

Acquisition is composed of subprocesses. The first is 'encoding', which in general refers the conversion of a message from one language, or code, to another. 'Encoding' is frequently used in the learning literature as a synonym for 'acquisition', but this is unsatisfactory, because there is more to 'acquisition' than 'encoding'. In neuronal encoding, information is transformed into the neuronal codes used in computation and representation (Churchland and Sejnowski 1992). This information arrives from either the external or the internal world. In the first case, the electromagnetic, mechanical, or chemical information is converted via the sense organs into neuronal activity. In the second case, information from the body itself is conveyed by specialized neuronal circuits, or via body fluids in the form of chemical messages (hormones) that evoke neuronal activity. No information can be handled by the central nervous system without first being encoded into the appropriate neuronal code. Encoding is thus involved in brain activities that do not necessarily culminate in the acquisition of a memory, such as on-line processing of information (*attention, *percept), or control of ongoing physiological routines. For a memory to be born, an additional process, of initial 'registration' ('recording'), is also needed. This permits the *internal representations of transient *stimuli, once formed, to become or induce an engram. From what we know from physiology and psychophysics, the decay time of transient representations is in the subsecond range (Dudai 1997b, see also 'encoding time' in Ganz 1975; *cell assembly, *percept, *phase). The registration mechanisms hence differentiate transitory from lasting internal representations, where 'lasting' is anything that is significantly longer than the aforementioned decay time.

How much time does acquisition require? This depends on the learning *paradigm and protocol. It is convenient to distinguish 'instant' from incremental ('repetitive', 'rote') acquisition. Instant acquisition refers to single-trial learning. This takes place in certain situations of intense aversive conditioning (*conditioned taste aversion, *fear conditioning); in some types of *imprinting; in the formation of *flashbulb memories; and probably in some other situations, in which acquisition curves have a step-function shape (e.g. *insight). In contrast, incremental acquisition refers to situations in which information accumulates over multiple experiences to construct the memory (Pavlov 1927; Skinner 1938; Hebb 1949; Dudai 1989). Gradual acquisition of *habits and *skills is such a case. The repetitive practice is expected to involve gradual modification of internal representations over hours, days, even months. But does incremental acquisition involve accumulative modifications that are restricted to the original representation formed at the beginning of training? This assumption might be naive. Internal representations are expected to form dynamic distributed networks (*cell assembly). Therefore, a more realistic view is that recurrent discrete events of acquisition and consolidation, that stem from each accumulative experience, alter existing internal representations that encode the information in question, but at the same time generate new representations and link them to the old ones (*palimpsest).

Ample data, supported by learning theory, indicate that whatever happens in acquisition, in terms of perceptual *cues and cognitive processes, determines not only the lifespan of the resulting memory, whether short or long (Craik and Lockhart 1972; Baddeley 1997), but also how efficiently will this memory be *retrieved in due time. Two influential concepts that reflect this notion will be mentioned here. One is the 'encoding-specificity principle' (Tulving 1983). It states that memory performance is best when

the *cues* present at retrieval match those present in acquisition. The other is termed '*transfer-appropriate processing' (Morris *et al.* 1977). It states that memory performance is best when the cognitive *processes* invoked at retrieval (say, semantic as opposed to phonetic processing in verbal tasks) match those used in acquisition.

Multiple approaches are used to investigate the neurobiology of acquisition. Cellular physiology, neuropharmacology, neurochemistry, and molecular biology are all applied to dissect the molecular and cellular mechanisms involved. Candidate 'cellular acquisition devices' are *ion channels and membrane *receptors on synaptic terminals that receive the teaching input, itself encoded in ion currents and *neurotransmitters (**Aplysia*, *long-term potentiation). A substantial amount of information is also available on the processes downstream from the synaptic membrane, that involve activation of *intracellular signal transduction cascades, and couple acquisition to consolidation. We even seem to start to understand in molecular terms why is it that in many learning situations, distributed training with intercalated intervals between repetitive acquisition trials, is more efficient than massed, continuous training, in which acquisition mechanisms are expected to function nonstop (*spaced training).

Brain areas and neuronal circuits that subserve acquisition have been identified in *habituation, *sensitization, *classical, and *instrumental conditioning in a variety of *simple or less-simple *systems (e.g. **Aplysia*, *classical conditioning, *conditioned taste aversion, **Drosophila*, *fear conditioning, *honeybee). In recent years, *functional neuroimaging has made a remarkable contribution to the identification of brain systems that subserve acquisition in the human brain (e.g. Nyberg *et al.* 1996; Fletcher *et al.* 1997; Tulving and Markowitsch 1997; Buckner and Koutstaal 1998; Epstein *et al.* 1999; Fernández *et al.* 1999). The circuits that acquire information about a memory vary with the type of memory, but a few general conclusions emerge from the studies so far: (a) acquisition of *declarative memories engages widely distributed areas, which include modality specific *cortex, and in addition supramodal areas, particularly in the mediotemporal lobe (*hippocampus, *limbic system); (b) these areas partially overlap brain areas that later retrieve the learned information; and (c) in some studies it was possible to show a correlation between the activation of an identified brain region during the training experience and the subsequent ability to remember this experience. For example, the ability to remember verbal information could be predicted by the magnitude of activation in the left prefrontal and temporal cortex during the training (Wagner *et al.* 1998b). It is not yet known, however, which of the activated areas is indispensable for acquisition (*criterion), which area is causally related to the strength of the engram, and what are the specific roles of each of the areas in the encoding and registration of information in the first milliseconds and seconds after engraphy has been triggered.

Selected associations: Consolidation, Experimental extinction, Retrieval, Transfer

Algorithm

A procedure for solving a problem or achieving a goal in a finite number of steps.

'Begin at the beginning', said the King of Hearts, 'and go on till you come to the end: then stop'. He thus provided White Rabbit with an algorithm (Carroll 1865). The term 'algorithm' is derived from Latinization of the name of one of the most creative mathematician in medieval Islam, Al-Kwarizmi (780–*c.* 850; Boyer 1989; Colish 1997). In modern times algorithmics is a field fundamental to the science of computing (Harel 1987). In the neurosciences algorithms are encountered in multiple contexts (Marr 1982; Hinton 1989; Churchland and Sejnowski 1992). One of these is in *models of biological learning. It is noteworthy that in discussion of such models the terms 'law', 'rule', and 'algorithm' are sometimes intermixed. It is therefore useful to distinguish among them. A 'law' is a scientifically proven formal statement with theoretical underpinning that describes a quantitative relationship between entities. Strictly speaking, there aren't yet *bona fide* 'laws' specific to the discipline of biological memory. It is sensible, therefore, not to misuse the term. 'Rule' describes a standard procedure for solving a class of problems. It is hence close to 'algorithm'. However, they are not equivalent. 'Algorithm' is a formal term referring to a detailed recipe, whereas 'rule' may be vaguer. Furthermore, a 'rule' may connote knowledge by the executing agent of the input–output relationship, 'algorithm' does not. A *system can execute algorithms perfectly without having the faintest idea what it is doing, why it is all done, and what the outcome is likely to be. As there is no *a priori reason to assume that biological learning at the *synaptic or circuit *level is governed by a knowledgeable supervisor (*homunculus), it does not make a lot of sense to claim

that synapses or circuits follow 'rules'; rather, they execute algorithms. Finally, an assumption (usually tacit) of the neuroscience of learning, and an incentive for the analysis of *simple systems, is that a great variety of biological learning systems, in different species, share general laws/rules/algorithms. This posit makes sense if evolution is considered, but is definitely not itself a law, and its generality must be scrutinized in every experimental system anew (e.g. Seligman 1970).

The most popular algorithms in the neuroscience are synaptic ones, and are associated with a postulate of synaptic *plasticity dubbed 'Hebb's postulate'. In its original version it states the following: 'When an axon of cell A is near enough to excite a 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' (Hebb 1949; for rudimentary precedents see James 1890; Kappers 1917). In a Hebbian synapse, the increase in synaptic weight is thus a function of the correlation of pre- and postsynaptic activity. Hebb postulated the process to account for experience-dependent modification of local nodes in *cell assemblies. In formal notation, Hebb's postulate is of the type $w_{i,j}(t+1) = w_{i,j}(t) + \Delta w_{i,j}(t)$, where $\Delta w_{i,j}(t) = f\,[a_i(t), a_j(t)]$; $w_{i,j}$ is the strength ('weight') of the connection from presynaptic unit u_j to postsynaptic unit u_i, $\Delta w_{i,j}(t)$ is the change in synaptic strength, $a_j(t)$ and $a_i(t)$ are measures of pre- and postsynaptic activity (Brown *et al.* 1990). Each step in the algorithm is thus a computation of the aforementioned type, and the algorithm consists of proceeding step-by-step over time (at a more *reduced level, the Hebbian computation itself is based on multiple subordinate algorithms, such as summation and multiplication, but this should not concern us here). The original 'Hebbian' became a generic term as well as a reference for many variants of synaptic modification algorithms. Terms composed of 'Hebb-plus-a-modifier' to mark their relationship to the Hebbian are common, and sometimes a bit confusing. For example, 'anti-Hebb' is used to describe rather different types of algorithms that culminate in decrement of synaptic efficacy (e.g. Lisman 1989; Bell *et al.* 1993; *long-term potentiation, *metaplasticity). Over the years multiple attempts have been made to demonstrate how Hebbian algorithms might be implemented in synapses in *development and learning (e.g. Lisman 1989; Fregnac and Shulz 1994; Buonomano and Merzenich 1998; Lechner and Byrne 1998; but see a critical review in Cruikshank and Weinberger 1996).

A discipline in which synaptic learning algorithms became particularly popular and useful is that of artificial neural networks (ANN; Fausett 1994). These are artificial systems (i.e. either abstract *models or the physical implementation of such models) composed of a large number of interconnected computational units ('neurons'). Signals are passed between neurons over connections, which manipulate the signal in a typical way. Each neuron applies an activation function to its net input to determine its output signal. Specific networks are characterized by the pattern of their connectivity ('architecture'), the algorithm that determines the weight on the connections, and the activation function of the neurons. The collective behaviours of such networks could mimic various dynamic properties of neuronal circuits, such as *perception and learning. Certain subclasses of ANN use Hebbian algorithms to achieve 'unsupervised' learning (see above) in local nodes. Other algorithms refer to 'supervised' learning, in which some type of global information or 'instructor' informs the node what the desired end-point is. An algorithm of the latter type that has gained considerable popularity is 'back-propagation' (or 'back-propagation of errors'). Here the error for each unit (the desired minus the actual output) is calculated at the output of the network, and recursively propagated backward into the network, so that ultimately, the weights of connections are adjusted to approach the desired output vector of the network (Rumelhart *et al.* 1986a).

A number of algorithms have been proposed to underlie learning at the more global levels of brain and behaviour (Thorndike 1911; Dickinson 1980; Wasserman and Miller 1997). An influential one is associated with the Rescorla and Wagner model of learning (1972; for precursors, see Hull 1943; Bush and Mosteller 1951). Basically, Rescorla and Wagner posited that in *associative learning, changing the associative strength of a stimulus with a *reinforcer, depends upon the concurrent associative strength of all present stimuli with that reinforcer; if in a given training trial the composite associative strength is already high, learning will be less effective. In formal notation, Rescorla–Wagner propose that $\Delta V_X = \alpha_X \beta_R (\lambda_R - V_\Sigma)$, where ΔV_X is the change produced by a given training trial in the strength of the association (V_X) between stimulus X, and reinforcer R; α_X and β_R are learning rate parameters (associability parameters) representing properties such as the intensity and saliency of X and R; λ_R is the maximal conditioning supportable by R; and V_Σ is the total associative strength with respect to R of all the stimuli present on the aforementioned trial. The expression $\lambda_R - V_\Sigma$ can be said to represent the disparity between expectation and reality on a given trial; the smaller it is, the weaker is the learning. In other words, as many

a reader might have concluded from their own experience, the amount of learning is proportional to the amount of *surprise (see also *attention). Here again, each step in the algorithm is a computation of the aforementioned type, and the algorithm consists of proceeding step-by-step over time. The Rescorla–Wagner model can explain multiple behavioural phenomena in conditioning, including cases of *cue revaluation (Dickinson 1980; Wasserman and Miller 1997; *classical conditioning).

Over the years multiple attempts have been made to account for the operation of selected brain regions by proposing identified synaptic and circuit algorithms (For notable examples, see Marr 1969; Albus 1971; Zipser and Andersen 1988). At the current state of the art in brain research, synapses and model circuits still provide a more suitable arena than whole real-life circuits to identify and test learning algorithms, because the input–output relationship of real-life brain circuits is seldom understood in reasonable detail, if at all. Still, advances are being made at more global levels of brain function as well; for example, Schultz *et al.* (1997) report that in the course of multitrial instrumental training, *dopaminergic activity in the primate brain encodes expectations and prediction errors for reward. The dopaminergic neuro-modulatory system may thus be part of a circuit that performs computations of the type $\lambda_R - V_\Sigma$ in the Rescorla–Wagner model.

New classes of algorithms are expected to emerge at the cellular, circuit, and system levels with the intensification of the mechanistic revolution in biology. One of these days, much of descriptive neurobiology is bound to give way to a science of biological engineering, in which algorithms and quantitative relations will become the rule rather than the exception. This has profound implications concerning the proper education of future neurobiologists (e.g. Alberts 1998).

Selected associations: Learning, Models, Level, Plasticity, Synapse

Amnesia

1. The loss or absence of memory.

2. The *amnestic syndrome*: A marked, chronic impairment in memory in the absence of other major cognitive deficits.

Amnesia is 'forgetfulness' in Greek (*mnemonics). The adverse effect of certain types of brain injury and mental trauma on memory was recognized long ago. But the systematic analysis of amnesia started only in the nineteenth century, with Ribot (1882) and Korsakoff (1887). Till the introduction of *functional neuroimaging, the study of amnesia has been the only practical approach to the investigation of brain substrates of memory in humans. Some information could be also obtained from electrical stimulation of patients undergoing brain surgery, but this was very limited in scope and controversial in interpretation (*engram). The investigation of amnesia is still a very powerful, unique approach to the analysis of human memory: whereas the application of functional neuroimaging could identify *correlations* between the activity of distinct brain regions and the *performance on memory tasks, the study of amnesiacs could potentially identify those brain structures that are *obligatory* for normal memory (*criterion, *method).[1]

Amnesia is not a unitary syndrome (Whitty and Zangwill 1966; Parkin 1987; Mayes 1995). A *taxonomy based on etiology distinguishes among 'organic amnesia', 'substance-induced amnesia', and 'functional amnesia'. These subtypes of amnesia are also known by other names, as explained below.

1. *Organic amnesia* is a consequence of damage to the brain inflicted by injury, disease (e.g. tumour, stroke, viral infection), or surgical intervention (DSM-IV 1994).
2. *Substance-induced amnesia* results from the intake of poisons, drugs of abuse, or medications with amnestic side-effects (for example, certain anxiolytics, *lotus). Chronic excessive consumption of alcohol could result in vitamin deficiency and encephalopathy (brain inflammation), which is manifested in Korsakoff's amnesia, at which stage it is also categorized as organic amnesia (Shimamura *et al.* 1988).
3. *Functional amnesia* develops after severe mental stress or trauma, or as a result of certain affective disorders. This type of amnesia is also termed 'psychogenic', or 'dissociative' ('dissociative disorders' in general are disruptions in the integrated functions of *consciousness, perception, personal identity, or memory).

The amnestic syndrome impairs learning and memory while leaving other cognitive faculties relatively intact. It is hence distinguished from *dementia, which involves multiple cognitive deficits, and from delirium, which impairs consciousness. Whereas some amnesia

are modality specific (e.g. Rubin and Greenberg 1998), the 'amnestic syndrome' is 'global' and independent of sensory modality. Global organic amnesia is *chronic*; some improvement may be observed over time, but the patient does not regain normal memory. There is also a separate syndrome termed '*transient* global amnesia'. This is a benign neurological syndrome in which the onset of amnesia is sudden and the recovery fast (usually < 1 day). Transient amnesia could also follow head trauma or electroconvulsive therapy.

An additional *criterion used to classify amnesia is the temporal window to which the memory loss refers. Here a distinction is made between '*retrograde*' and '*anterograde*' amnesia. Retrograde (premorbid) amnesia affects memory from the onset of the pathology backward. Anterograde (postmorbid) amnesia affects memory from the onset of the pathology forward. For example, in a typical case of the amnestic syndrome, there is dense anterograde amnesia and usually only a partial, graded retrograde amnesia. Memory of the recent past is commonly affected more than memory of the distant past; this observation is termed 'the law of regression', or 'Ribot's law' (it is noteworthy that Ribot regarded the phenomenon as the manifestation of a Darwinian principle, in which 'progressive destruction advances progressively from the unstable to the stable'; Ribot 1882).

The *classical, most widely cited case of a global amnesia is that of H.M. He became amnestic in 1953 at the age of 23, following 'a frankly experimental operation' (Scoville and Milner 1957) to alleviate uncontrollable epilepsy. The operation removed bilaterally the medial temporal polar *cortex, most of the *amygdaloid complex, the entorhinal cortex, and approximately half of the rostrocaudal extent of the intraventricular portion of the *hippocampal formation (Corkin *et al.* 1997). The operation reduced the frequency of seizures, but produced a severe, permanent anterograde amnesia, with only a limited effect on memory of events prior to the operation (and no effect on more remote events). Postoperationally, H.M. scored above average on a general intelligence test, showed no decline on immediate memory (*capacity), but was unable to store any new *declarative information. He was, however, capable of learning new *skills. Thus even in this severe case, the amnesia was not really 'global'.

The study of H.M., as well as of many other amnesics since then, gave rise to major insights concerning human memory (Squire and Zola 1997; Milner *et al.* 1998). These studies have demonstrated that the brain

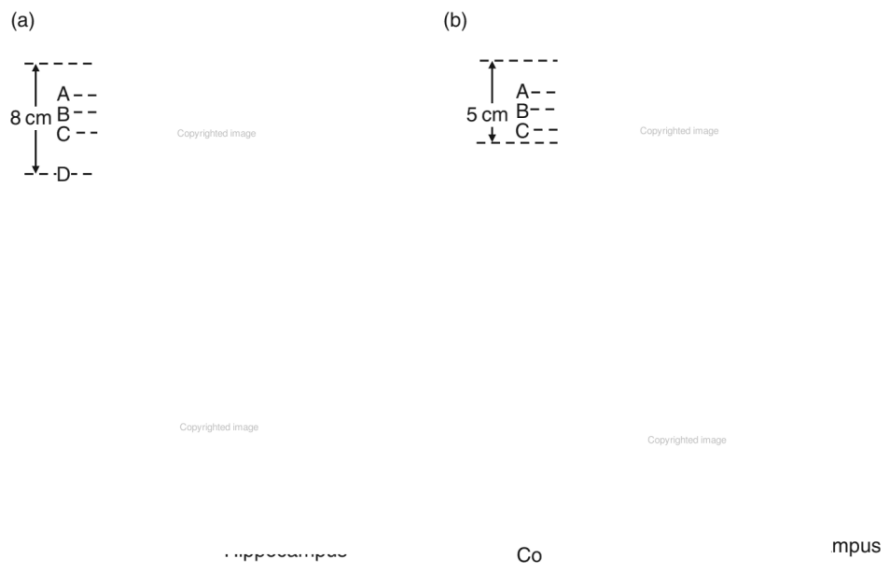

Fig. 2 The missing parts in the brain of H.M., removed in the operation that had resulted in global amnesia. (a) The surgeon's estimate after the surgery (Scoville and Milner 1957). (b) The outcome of the surgical resection as unveiled by magnetic resonance imaging (MRI) 40 years later (Corkin *et al.* 1997). The upper diagrams depict ventral views of the brain, the lower ones depict coronal sections. A through D in the ventral views mark the planes of coronal sections in the original drawings, but only plane B is shown here. The operation was bilateral but in the drawing one hemisphere is shown intact for comparison. Adapted from Corkin *et al.* (1997). The case of H.M. drew much attention to the role of the medial temporal lobe in general, and the hippocampus in particular, in long-term memory.

contains distinct declarative (explicit) and nondeclarative (implicit) memory systems; and that long-term declarative memory is dependent on medial temporal lobe structures. Additional research has shown that nondeclarative amnesia could result from damage to a different, corticostriatal system (Mishkin *et al.* 1984; Knowlton *et al.* 1996; *skill). Support for the above conclusions has also emerged from studies of circumscribed brain lesions in *monkey *models of human amnesia (e.g. Mishkin *et al.* 1984; Ridley and Baker 1991; Meunier *et al.* 1993; Zola-Morgan *et al.* 1993; Gaffan 1994; Leonard *et al.* 1995). Indeed, the neuroscience of amnesia is characterized by a remarkable degree of integration of human and animal research.

Despite the impressive advances in our understanding of amnesia, many outstanding questions still await resolution (Warrington and Weiskrantz 1982; Mishkin *et al.* 1997; Nadel and Moscovitch 1997; Squire and Zola 1997; Weiskrantz 1997; Milner *et al.* 1998; Aggleton and Brown 1999). Among these: Is amnesia due to impairment in the *acquisition, *consolidation, storage, or *retrieval of memory? Although most authorities consider acquisition of information to remain intact in global amnesics, because of the good performance on the immediate memory tasks (see H.M. above), still, even subtle deficits in the way information is encoded and registered could markedly affect later retrieval. Another question is what is the specific contribution of medial temporal lobe structures (such as the hippocampal formation and adjacent cortici), and medial diencephalic structures (such as the medial thalamus and the mammillary bodies), to different manifestations of the amnestic syndrome, such as anterograde vs. retrograde amnesia, or *recall vs. *recognition deficits? And what is the contribution to amnesia of other brain areas, such as the basal forebrain (*acetylcholine, *dementia), or the frontal cortex and its interconnections with the diencephalon?

Each amnestic *subject is a unique individual, and probably in none are the lesions confined to a single well-circumscribed functional location in the brain. This makes the research inherently difficult. Animal models do help a lot, but still, it must be proven that what is considered amnesia in a monkey, even more so in a rodent, is sufficiently similar to the human amnesia to warrant adaptation of the conclusions from the animal to the human. Solutions are expected to emerge from the systematic analysis of additional cases of amnesia (e.g. Reed and Squire 1998), using universally accepted batteries of memory tests; from a greater sophistication of such tests in humans, primates, and rodents; and possibly also from a more extensive integration of novel functional neuroimaging methods in the study of amnestic brains.

Selected associaions: Conscious awareness, Declarative memory, Episodic memory, Dementia, Infantile amnesia

[1] Reversible disruption of activity by transcranial magnetic stimulation (TMS) might also be used to identify brain areas obligatory for learning and memory (e.g. Grafman *et al.* 1999; Rossi *et al.* 2001), but it has not yet been widely employed.

Amygdala

A heterogeneous collection of nuclei and cortical areas in the temporal lobe, considered to subserve emotional and social behaviour, learning, and memory.

The amygdala (alias the amygdaloid or amygdalar complex), first described and named by the German anatomist Burdach in the early nineteenth century (Meyer 1971), is so called because in the primate brain its shape resembles an almond (*amugdalē* in Greek). About a dozen different nuclei and specialized cortical areas are currently discerned in the amygdala, and many intra- and extra-amygdalar connections have been identified (Amaral *et al.* 1992; Pitkanen *et al.* 1997; Swanson and Petrovich 1998; Aggleton 2000). Indeed, the heterogeneity of the nuclei, areas, and pathways raised some doubts whether 'amygdala' as a whole is a discrete anatomical entity *in situ*, or only an artificial construct of the human mind (e.g. Kirkpatrick 1996; Swanson and Petrovich 1998; de Olmos and Heimer 1999). Whether a well-defined natural kind or merely a convenient concept, judging by its connectivity, the amygdaloid complex fits well to serve as a central processor for some facets of sensory and supramodal *representations. This is because sets of amygdaloid nuclei interconnect heavily with the unimodal and polymodal *cortex, as well as with subcortical structures. Some of these pathways are asymmetrical (more extensive in one direction, e.g. from amygdala to hippocampus), and the information flows into one amygdaloid nucleus but comes out at another.

The peculiar behavioural effects of bilateral lesions of the temporal lobe, including the amygdala, were noted over a century ago in *monkeys (Brown and Schafer 1888), and later further characterized (Kluver and Bucy 1938) and termed the 'Kluver–Bucy syndrome'. The overall impression was that the lesion

produced 'a condition resembling idiocy' (Brown and Schafer 1888). A more detailed look described the lesioned animals as tamed, over-attentive but fearless, devoid of the ability to assess the significance of inanimate and animate objects, and indiscriminately phagic and sexual. A similar syndrome was shown to result from ablations confined to the amygdaloid complex and the medial temporal polar cortex (Weiskrantz 1956). It is indeed likely that many functions used to be attributed to the so-called *'limbic system', including control of phylogenetically primitive drives, emotions, and elementary social interactions, are carried out by the amygdala (LeDoux 1991).

Over the years, circumscribed lesions in monkeys and rodents, cases of diseased and injured amygdala in humans, and recently *functional neuroimaging, have all been employed to investigate the role of the amygdala in learning and behaviour. The effect of amygdala dysfunction on a number of *recognition tasks, including *delay tasks and visual and cross-modal associations, was first taken to imply that the amygdala plays a major part in these tasks; however, later studies indicated that the impairment was due to damage to the adjacent rhinal cortex, which was injured together with the amygdala in the original lesion experiments (Zola-Morgan *et al.* 1989*a*,*b*; Murray *et al.* 1993). In contrast, conclusive evidence for the involvement of amygdala in learning and memory was found in other types of tasks, which engage fear and emotional memory (Adolphs *et al.* 1995; Maren and Fanslow 1996; Rogan and LeDoux 1996; Scott *et al.* 1997; Walker and Davis 1997; Cahill and McGaugh 1998; Lamprecht and Dudai 2000; Parkinson *et al.* 2000). A most popular paradigm in this context is Pavlovian *fear conditioning, a ubiquitous form of *classical conditioning. In Pavlovian fear conditioning, a conditioned stimulus (e.g. tone) is associated with an aversive unconditioned stimulus (e.g. electric shock), to yield fear (e.g. freezing, increased blood pressure and heart rate) as the conditioned response. Amygdalar nuclei, including a subset dubbed the 'amygdalar basolateral complex', were specifically implicated in this simple type of conditioning (the identity of nuclei recruited in fear conditioning is probably also a function of the task complexity; Killcross *et al.* 1997; Nader and LeDoux 1997).

The meticulous analysis of fear conditioning in the amygdala had clearly paid off: it has yielded the first demonstration of *long-term potentiation induced by training in an identified pathway that subserves learning in the behaving rat (Rogan *et al.* 1997). The cellular analysis of fear conditioning also strengthened the assumption that the amygdala itself is a structure that stores information (see also Lamprecht *et al.* 1997). A different view is that the amygdala, occupying a strategic position in the network of widespread neuromodulatory systems in the brain, does not itself store memory, but rather modulates other circuits that store it (Cahill and McGaugh 1998). The clash between these opposing views has raised central issues concerning memory traces: is the evidence for the requirement for *protein synthesis and gene expression in training sufficient to prove that a certain brain area *consolidates a given memory? And if it is, will the memory be stored in that area forever after? And which parts of a circuit that subserves a memory should be considered as an integral part of the postulated *engram? On top of it all, there is actually no reason to assume that the 'storage' and the 'modulation' views are mutually exclusive. Moreover, even a close look at the Kluver–Bucy syndrome indicates that there is more to the amygdala than storage, and that it regulates *attention and additional facets of cognition (Gallagher and Holland 1994).

The study of the role of amygdala in fear conditioning is a beautiful example of a cross-*level analysis that has led from the behaving organism to circuits, *synapses, and molecules, and *vice versa*. An issue that deserves further emphasis is the ethological context of the findings. The amygdala fulfils an important role in navigating the individual in its species-specific milieu, enabling it to construe sign-*stimuli correctly, and react

Fig. 3 The amygdaloid complex maintains extensive interconnections with multiple brain areas, including the hypothalamus, thalamus (MD, mediodorsal), *hippocampal formation, and temporal and frontal *cortex. This schematic diagram depicts the amygdala as a single area for simplicity, but in reality it is a collection of about a dozen main nuclei and cortical areas that interconnect differentially with targets over widely distributed brain areas, and subserve diverse functions, among them emotional behaviour and learning. (Adapted from Brodal 1998.)

to them appropriately (e.g. see the role of amygdala in perception, memory, and judgement of facial as well as verbal expression in humans, Adolphs *et al.* 1998; Morris *et al.* 1998; Isenberg *et al.* 1999). This is definitely a place to look for brain defects that underlie some neurotic and affective disorders and asocial behaviours.

Selected associations: Fear conditioning, Functional neuroimaging, Limbic system, Long-term potentiation

Anthropomorphism

The attribution of human attributes to mythical creatures, inanimate objects, or nonhuman organisms.

The term is derived from Greek: *ánthropōs*—human being, *morphé*—form. Anthropomorphism owes much to anthropocentricity, i.e. our *a priori inclination to regard ourselves as the centre of the universe and see the world through our *biased eyes. By doing so we probably hope to gain some illusory control over reality. Anthropomorphism is intensively and recurrently exemplified in ancient myths, literature, and art (e.g. Burkert 1985). Occasionally, it had also infiltrated other social activities: throughout Europe in the Middle Ages, horses and pigs were dragged to public trial because it was believed that they are *consciously aware of their own acts and hence are liable for them (Evans 1906). In the early days of experimental psychology, anthropomorphism was popular (Boakes 1984), being influenced by the Darwinian theory of evolution that suggested a mental continuum along with the physical one. The *classics of the anthropomorphic tradition in animal psychology are books by Darwin (1872) and Romanes (1882). The transformation of psychology into a more objective and quantitative scientific enterprise was accompanied by attempts to abandon anthropomorphic anecdotes that portrayed pets as geniuses, and to adhere to parsimonious explanations of animal behaviour, such as advocated by Loyd Morgan's canon (*Ockham's razor). However, anthropomorphism still pops out between, and occasionally in, the lines of current research articles in biology and psychology (e.g. see discussions in Kennedy 1992; Sullivan 1995).

Anthropomorphic accounts could be classified into two kinds: *metaphorical and explanatory. The metaphorical are the more innocent ones. They may add colour to an otherwise rather dry scientific account. To describe the behaviour of protozoa as '… if they did not enjoy being alone and had passed the word along to gather and hold a mass meeting' (Jennings 1899) is a matter of style only, as far as the description does not lead the reader (and even more so the writer) to assign to the unicellular organism *declarative human-like social drives. Explanatory anthropomorphism, however, may result in embarrassing errors. A trivial example is the exposure of teeth in monkeys; what could be construed by the approaching novice as a friendly smile might actually be an expression of threat.

Possibly most relevant to current neurobiological research is our innate tendency for implicit anthropomorphism, i.e. tacitly construing the behaviour of animals in terms of problem solving *algorithms that could have been used by the human observer. This should especially be taken into account in cases in which sophisticated cognitive faculties are suggested, for example the formation of cognitive *maps in insects (Wehner and Menzel 1990), of *learning sets in rodents (Reid and Morris 1993), or of *observational learning in invertebrates (Fiorito and Scotto 1992). Implicit anthropomorphism may result not only in superfluously complex explanations but also in excessively austere ones. As these lines are being written, hundreds of diligent postdocs are running rats or mice in water *mazes, assuming that from the outset, all that the wet animal has in mind from the outset of the experiment is the urge to learn the shortest way to the platform and take a break, because this is what the experimenter would have done. While still wishing to escape the water, in reality, some of the drives and strategies pursued by the swimming rodent are species specific (e.g. Wolfer *et al.* 1998).

There is, however, a twist to the story. In spite of the aforementioned caveats and reservations, the mere fact that an explanation has an anthropomorphic connotation is not sufficient to demote it. In other words, 'anthropomorphism' *per se* cannot be used as a *criterion in refuting or accepting explanations and *models. The truth is that we do not really know the borders between the mental faculties of other mammals and those that are sometimes considered as exclusive privileges of **Homo sapiens*. For example, when rodents associate events, are they *consciously aware of it (Clark and Squire 1998; *declarative memory)? And if they are, what is the depth and quality of their conscious awareness? In recent years, the more we learn about the physiology and psychophysics of animals, the more we become astonished to discover that even species far

remote from us on the phylogenetic scale seem to perceive some aspects of the world not so differently from us (e.g. Nieder and Wagner 1999). This raises the possibility that underestimating the capabilities of their brain is as misleading as overestimating it. There is no reason why we should not expect to find in evolution a gradient of antropolikeness on a great variety of faculties, such as *planning, *prospective memory, complex problem solving, or *insight. It is even still debated whether symbolic language had really emerged in humans only (Walker 1983; Griffin 1984; Cheney and Seyfarth 1990).

But, whereas some anthropolike mental faculties, such as numerical competence, are amenable to objective measurement (Davis and Perusse 1988; Brannon and Terrace 1998; Kawai and Matsuzawa 2000), others, e.g. subtle emotions, are not. We may therefore never be able to really know what it is like to be a bat (Nagel 1974). We are hence left with the humble conclusion that the interplay between prudent adherence to Ockham's razor on the one hand, and proper appreciation of the phylogenetic and ecological specialization of other species' brains on the other, is delicate indeed.

Selected associations: Artefact, Bias, Clever Hans, Declarative memory, Subject

Fig. 4 It works both ways: an apeomorphized version of Charles Darwin in a contemporary caricature. Faithful to the *zeitgeist that his theory of evolution reinforced, Darwin himself anthropomorphized animal behaviour in *The expression of the emotions in man and animals* (1872).

Aplysia

The sea-hare, a marine snail.

Aplysia, a hind-gilled (opistobranch) marine snail (Kandel 1979), is one of the heroes of the cellular revolution in the neurosciences. Its external resemblance to the rabbit earned it the name sea-hare. Yet it is the insides of *Aplysia* that has turned it into such a highly successful *system in the cellular analysis of simple memory. Quinn (pers. comm.) had defined an ideal *subject for the neurobiological analysis of learning as a creature with 10 large neurons, 10 genes, a generation time of 1 week, and the ability to play the cello and recite Shakespeare. Indeed this is not a faithful description of *Aplysia*, but in the real world, the sea-hare became a useful compromise. Its main assets are a relatively simple nervous system that is readily accessible to experimentation, a simple behavioural repertoire, and a group of capable investigators that have become fascinated by the virtues of the slug.[1]

The central nervous system of *Aplysia* is composed of about 20 000 nerve cells arranged in widely spaced ganglia (masses of nerve cells). Some secretory neurons are as big as the entire brain of **Drosophila*. Some neurons can be identified from one individual to another by their location, shape, and firing pattern. The system had attracted cellular physiologists (Arvanitaki and Chalazonitis 1958; Tauc and Gershenfeld 1961; Kandel and Tauc 1965). It was, however, the research on *plasticity and learning that has endowed *Aplysia*, especially *Aplysia californica*, with its fame (Kandel and Schwartz 1982; Byrne and Kandel 1996). Following a series of reductive and simplifying steps (*reduction), the cellular and molecular mechanisms of learning in *Aplysia* have been pursued from the behaving animal, via preparations of isolated ganglia, to identified nerve cells and *synapses in culture (Carew *et al.* 1971; Rayport and Schacher 1986; Bartsch *et al.* 1995; Frost *et al.* 1997; Hawkins *et al.* 1998). This system is the epitome of the reductionist approach to memory, and as such demonstrates both the advantages and the shortcomings of the approach.

Like all organisms with a nervous system, *Aplysia* display a repertoire of defensive (e.g. withdrawal) and appetitive (e.g. feeding) reflexes. The analysis of learning in *Aplysia* has focused mainly on the defensive reflexes (Kandel 1976; Byrne 1985). These can be illustrated by the gill-and-siphon withdrawal reflex (GSWR). The gill is the external respiratory organ of *Aplysia*. It is housed in the mantle cavity on the dorsal side of the animal. The cavity is a respiratory chamber covered by the

mantle shelf. At its posterior end, the shelf forms a fleshy spout, called the siphon. The siphon protrudes out of the mantle cavity between wing-like extensions of the body wall, called parapodia. If a tactile *stimulus is applied to the siphon or mantle shelf, a two-component reflex is elicited. One component is contraction of the siphon and its withdrawal behind the parapodia. The other is contraction of the gill and its withdrawal into the mantle cavity. The GSWR can be *habituated by repetitive monotonous tactile stimuli to the skin; *sensitized by noxious stimuli to the tail or head; and undergo *classical conditioning. This is achieved by pairing a gentle stimulus to the siphon or gill (the conditioned stimulus) with a noxious stimulus to the tail or head (the unconditioned stimulus), so that the conditioned stimulus comes to evoke intense withdrawal (the conditioned response).

In intact *Aplysia* the GSWR is controlled by both the central and the peripheral nervous systems. Most of the cellular analysis of learning has been performed in the central nervous system, particularly in the abdominal ganglion. This ganglion was found to subserve a substantial portion of the habituation, sensitization, and classical conditioning of the GSWR. Multiple sites of plasticity have been identified in the abdominal ganglion, but the attention has been focused primarily on one site: the synapse between the sensory neurons and the gill or siphon motor neurons (Kandel and Schwartz 1982; Byrne and Kandel 1996; Figure 5). It has been proposed that part of the behavioural plasticity of the GSWR could be accounted for by use-dependent modifications in this synapse. In brief, the cellular analogue of habituation was portrayed as presynaptic depression, induced by repetitive monotonous firing. As this depression involves only the modified synapse, it is said to be 'homosynaptic'. Sensitization was portrayed as synaptic facilitation, induced in the presynaptic terminal of the aforementioned sensory-to-motor

Fig. 5 A highly simplified scheme of a fragment of the circuit that subserves the gill-withdrawal reflex and its modification by experience in *Aplysia*. The reflex could be elicited by a tactile stimulus applied to the siphon skin. Repetitive, monotonous tactile stimuli result in habituation of the reflex. A shock to the tail results in sensitization of the reflex. Classical conditioning is obtained by pairing the shock to the tail with a light tactile stimulus to the siphon, so that this tactile stimulus comes to evoke intense withdrawal on subsequent applications in the absence of the shock. Probably hundreds of nerve cells and thousands of synapses subserve the reflex in the intact animal; only a selection of types of cells and synapses are depicted in the scheme. IN, interneuron; MN, motor neuron; SN, sensory neuron. The presynaptic terminal of the sensory-to-motor synapse, denoted by a black triangle (left-hand side), was so far the focus of much of the cellular and molecular analysis of the reflex. *Plasticity of this synapse contributes both to the short- and to the long-term *phases of memory in the reflex. For further details see text.

synapse by *neurotransmitters that are released from interneurons and encode the sensitizing stimulus (Figure 5). As this facilitation involves multiple types of synapses, it is 'heterosynaptic'. Classical conditioning of the GSWR was portrayed as sharing cellular mechanisms with sensitization. It is also activity-dependent presynaptic facilitation; however, in contrast with sensitization, which enhances the responsiveness to subsequent stimulation of the skin at any location, the facilitation in classical conditioning is specific to the pathway that has mediated the conditioned input (*coincidence detection). This is hence a pathway-specific, activity-dependent presynaptic facilitation. Multiple molecular mechanisms have been suggested to account for the *acquisition and short-term retention of the synaptic facilitation. They include activation of *intracellular signal transduction cascades by the facilitatory neurotransmitter(s), phosphorylation (by *protein kinases) of synaptic proteins (e.g. *ion channels), and modulation of transmitter release (Kandel and Schwartz 1982). These simplified cellular *models were later extended, enriched, and modified to include additional synaptic sites and mechanisms (e.g. Byrne and Kandel 1996).

Because of lack of space, we will not concern ourselves here with the fine details of the *Aplysia* story, but rather with a few generalizations only. The cellular analysis of *Aplysia* reflexes has shown that a significant component of the circuit that subserves simple learning could be pinned down to the *level of identified neurons and synapses. This analysis was the first to demonstrate the central role of cyclic adenosine monophosphate in memory (Cedar *et al.* 1972; *CREB), and the multiplicity of time- and *context-dependent mechanisms of plasticity in a single cell. It has also demonstrated that at least part of the loci that subserve short-term memory also subserve long-term memory. Further, analysis of plasticity in the GSWR has provided much support for the *zeitgeist proposal that long-term memory storage relies on modulation of gene expression (Goelet *et al.* 1986; Martin *et al.* 1997a,b; *consolidation, *immediate early genes, *protein synthesis). It is noteworthy that in recent years, much of the analysis of learning in *Aplysia* has practically merged with the cellular biology of *development. This may reflect a genuine homology between learning and development. Yet the focus on molecular and cellular mechanisms, which are shared with other disciplines in the life sciences, may also attest to the current difficulty in switching, even in a *simple system, to the more global level of analysis, which is critical for understanding memory, i.e. that of concerted circuit activity that ultimately encodes *internal representations in the behaving organism.

Over the years, the appreciation of the complexity of the *Aplysia* system has increased, and the highly simplified models gradually matured into more realistic ones (Glanzman 1995; Byrne and Kandel 1996; Fischer *et al.* 1997; Bao *et al.* 1998; Lechner and Byrne 1998; Royer *et al.* 2000). Attempts are also being made to elucidate the cellular bases of apetitive reflexes (Lechner *et al.* 2000), as well as of a more complex form of learning, *instrumental conditioning (Nargeot *et al.* 1999). *Aplysia* is still our main source of information about the molecular changes that take place in neurons up to a few days after training (*long-term potentiation addresses a shorter time window). This is evident among others from the references made to it in many entries in this book. Admittedly, the memory feats of *Aplysia* are modest (even the classical conditioning of the GSWR is only of the α type, namely, modification of a pre-existing behaviour and not acquisition of a novel one). But no doubt, without the remarkable work on *Aplysia*, the molecular and cellular biology of neuronal plasticity, learning, and memory would have been much, much duller. There is still one take home message that is worth mentioning here. The analysis of neuronal plasticity in *Aplysia* has unveiled an impressive inter- and intra-cellular molecular complexity that keeps growing. This should be noted by orthodox reductionists, who erroneously think that reducing a system implies simplifying it. The opposite might be the case.

Selected associations: CREB, Reduction, Simple system, Synapse

[1]The major driving force behind the *Aplysia* project, Eric Kandel, shared the 2000 Nobel prize for Medicine.

Artefact

1. Man-made object.

2. A phenomenon, process, or mechanism that does not normally exist in nature but is introduced by experimental manipulation of the *system.

3. A phenomenon, process, or mechanism that does not exist in nature but is believed to exist, due to erroneous interpretation of data or theories.

Artefact stems from the Latin 'something made with skill', but occasionally, in science, the major skill at stake is how to distinguish an artefact from a natural

phenomenon. Artefacts have haunted the experimental sciences since the emergence of the latter, much before the term was introduced into English at the beginning of the nineteenth century. In biology, 'artefact' was first used to denote aberrations produced in histological specimens by the fixation methods used to prepare the tissue for microscopic examination. However, with time, it came to embrace many types and tokens of artificial constructs,[1] either concrete or conceptual, which are confused with the real thing.

It is useful to distinguish two major classes of artefacts: technical (definitions 1,2) and conceptual, or interpretational (definitions 1,3). A harsh fixative or an unreliable stain leading to the appearance of an imaginary brain structure could be the cause of technical artefacts. Similar illusions may result from non-specific antibodies in an immunoblot, sloppy development of an autoradiogram, or tricky electrophysiological setups with a will of their own. Expert scientists come to master and prune the potential sources of artefacts in their trade, but new *methods and techniques generate new artefacts. For example, with more and more data analysis being relegated to fancy computer systems, the computers themselves become a source of technical artefacts before the data even reach the scientist. It takes a careful team leader to identify the problem (e.g. Katz *et al.* 1998).

A common potential source of interpretational artefacts is the so-called *post hoc* argumentation (*post hoc ergo propter hoc*, Latin for 'after this hence because of this'). *Post hoc* means arguing that because one event was correlated later in time with another, the second happened because of the first. This could sometimes be straightened out by performing *control experiments in which the order of events is altered or the suspected cause omitted from the protocol. For example, suppose we are tempted to conclude that a *receptor for the *neurotransmitter *glutamate in the *rat *hippocampus is phosphorylated (*protein kinase) as a consequence of learning to navigate in a *maze, because the receptor molecule appears phosphorylated after the experience; this might be a *post-hoc* artefact rather than a real consequence of the learning experience (e.g. see *criterion).

Interpretational artefacts could also result from lack of expertise in, or awareness of, a domain of knowledge that is relevant to the finding. This is a risk encountered especially, but definitely not solely, by investigators who shift from one field to another. A study of conditioning illustrates the case. In the first half of the last century, many operant conditioning paradigms ignored the species-specific behavioural repertoires of the experimental animals. This led to questionable conclusions. Probably hundreds of Ph.D. theses interpreted the pecking of pigeons in a Skinner box as an *instrumentally conditioned response; however, pecking is an innate response, the pigeons emit it anyway, and the situation might not have been instrumental but rather *classical conditioning (Jenkins and Moore 1973). Similarly, over the years, cats were reported to be meticulously conditioned to emit stereotypic behaviours in order to escape from puzzle boxes; but some of the typical behaviours, such as rubbing the flank or head against a pole, were later pointed out as species-specific feline greeting reactions, emitted in response to the observer rather than conditioned by the escape (Moore and Stuttard 1979). Note that here the artefact is both technical (due to the improper design of the experiment, allowing the observer to affect the behavioural response of the *subject), and interpretational. The role of the observer is probably continued to be ignored to this day in many labs; it would be of interest to enquire how often an unexpected behaviour of a rat in a *maze reflects an artefact due to the introduction into the room of a new perfume or after-shave or an admiring visitor. *Anthropomorphism is another potential source of interpretational artefacts, confusing innate (*a priori) species-specific behaviours with higher-order cognitive faculties (*clever Hans, *Ockham's razor).

Interpretational artefacts could also be due to variables unknown in the discipline at the time that the interpretation is being attempted. An example is provided by a study of the effect of exploratory behaviour on hippocampal neurons. When rats are transferred to an unfamiliar environment they explore and learn it. It has been reported that such exploration is accompanied by hippocampal *plasticity, including persistent facilitation of evoked neuronal responses (Green *et al.* 1990). Although the basic finding was confirmed (Moser *et al.* 1994), it later became clear that the effect is much smaller than first reported. The reason: fluctuations of up to 2–3°C in brain temperature, occurring during the exploratory activity, modify neuronal properties *in vivo* and account for a substantial part of the observed 'plasticity' (Andersen and Moser 1995). The aforementioned temperature effect and the resulting potential artefacts were not recognized at the time. In this case, the artefact was not a waste of intellect; its exploration led to new insights on the tricks of brain physiology. In other words, it is absolutely possible to learn even from artefacts.

Whether the suspected artefact is of the technical or the interpretational type, the first-law-of-the-artefact frequently holds: the more important is the message, the faster is the artefact exposed. Artefacts that lead

to boring conclusions gain immortality in obscure journals. But if the news is smashing, for example, that specific memories can be transferred from one individual to another in brain extracts (Babich *et al.* 1965; Ungar and Oceguera-Navarro 1965), the scientific community does its best to sort the facts out, even if the causes of the artefact, or at least what appears to be an artefact to the contemporary eye, do not always become clear in the process (Byrne *et al.* 1966; Nicholls *et al.* 1967; Smalheiser *et al.* 2001).

Selected associations: Anthropomorphism, Control, Red herring, Scoopophobia

[1]On 'types' and 'tokens', see *system.

Assay

A procedure or technique for the analysis of a phenomenon, process, or mechanism; a test.

Assays (from *exagiere*, Latin for 'to weigh out') are not merely research tools. They play a decisive part in the development and workings of scientific disciplines. They are also important in shaping the feasibility, progression, and outcome of particular research programmes. Sometimes they even play a decisive part in moulding the fate of individual academic careers.

Scientific assays are the nuts and bolts of scientific *methods and *paradigms. In experimental science, they are the end instrument used to embody the objectives of a 'method' and test the concepts of a 'paradigm'. They are thus more specific than 'methods'. The same method may be implemented by using a variety of assays. For example, one could employ a correlative *method* to probe the role of an *immediate early gene in memory in a given brain region, but use different *assays* to determine whether the expression of that gene is correlated with the behavioural change. A useful assay yields results that are then subjected to analysis and construed according to selected *criteria.

The spectrum of assays used in the neurosciences is rich and heterogeneous. Practically all these assays could also be incorporated into research programmes that target learning and memory. One useful classification of assays (*taxonomy) is by the *level of analysis involved. Other classifications are of course possible; for example, by the method that guides and utilizes the assay (i.e. correlation, intervention, etc.). Straightforward classification by level is into molecular, cellular, neuroanatomical, *system, and behavioural. In considering levels of analysis, one should note differences in the dialects of the scientific *culture. The term 'assay' is mostly popular in molecular and cellular studies. Neuroanatomists prefer to use 'technique' or 'method' (which, as noted above, is better reserved for a more comprehensive activity). Psychologists cling to 'test'. The latter term commonly carries the connotation of 'success' or 'failure' in *performance; 'assay' does not. 'Test' can also be used to denote particular instantiation of a type of assay in an experimental protocol.

Molecular assays, such as binding of drugs to *receptors or measuring enzyme activity, are shared by many branches of molecular and cellular biology (R. Martin 1997; e.g. *development). Cellular, neuroanatomical, and system procedures are shared by many subdisciplines of the neurosciences and are not unique to the study of plasticity and memory. One notable exception that comes to mind is *long-term potentiation, which under certain circumstances may be regarded as an assay to determine induction and maintenance of cellular *plasticity, although it is also a method, and moreover, a *paradigm. In contrast to molecular, cellular, and system assays, behavioural assays used in the field of learning and memory are unique to this field: they are specific 'memory assays' or 'tests'.

Some memory tests were groundbreaking at the time of their introduction. For a field of knowledge to become a scientific discipline, research techniques and assays are required that permit quantification of phenomena addressed in that field: '…the forces and actions of bodies are circumscribed and measured either by spatial intervals, or by moments of time, or by concentration of quantity, or by predominance of power; and unless these four are accurately and carefully weighed, the sciences concerned will be elegant speculations perhaps but of no practical use' (Bacon 1620). A handful of tests, by the mere fact that they had enabled for the first time the quantification of memory, had transformed the study of memory into a science. A prominent example is provided by tests involving *recall of series of so-called 'nonsense' syllables, introduced by Ebbinghaus to measure *forgetting (1885; see also Jacobs 1887). This type of experiment is considered to have opened the scientific era in research of human memory. Similarly, introduction of *classical and *instrumental conditioning has permitted the systematic experimental investigation of animal learning (Thorndike 1911; Pavlov 1927; for more on the history, see Boakes 1984).

Still another class of assays includes those that alter and reroute the course of a discipline. Here are some examples: Introduction of the *maze (Small 1901;

*classic) has paved the way to research on spatial learning, cognitive *maps, and other facets of memory. A popular descendent of those original mazes is the extensively used water maze (Morris 1981). Introduction of the *delay task (Hunter 1913) has permitted analysis of *recognition and *working memory, and development of *monkey models of *amnesia. Very useful versions are the trial-unique delay tasks, such as trial-unique delayed non-matching-to-sample (Gaffan 1974; Mishkin and Delacour 1975; *delay task). In some cases, adaptation of a well-known type of memory assay to a new organism could open a whole new field. An example is provided by olfactory conditioning in the fruit fly, **Drosophila*. Sophisticated *neurogenetic analysis of memory became feasible only after classical conditioning had been adapted to the special needs of the fly (Quinn *et al.* 1974). And, of course, there are those many assays that are variations on a theme, introducing important improvements and modifications to already existing methods.

Lack of an appropriate assay may hinder the development of a field or the resolution of a major research problem. For example, some types of behavioural assays engage the *hippocampus and are sensitive to hippocampal damage. However, at the time of writing there is still no satisfying behavioural assay to tap exclusively into hippocampal function in primates. Such a task will be very useful in clarifying the role of the hippocampus in memory. The hippocampus can also be invoked to illustrate a potential problem in the use of assays. This is the problem of 'circular argumentation'. Thus, given that a hippocampal lesion impairs performance on task X under condition A, some investigators are quick to use task X under conditions other than A to determine whether the hippocampus is involved, as if task X is an established probe for hippocampal involvement. Failure or success on task X, however, may result from parameters specific to condition A that do not *generalize to other conditions of the *subject or the experiment. The problematics are further augmented when inference is made from one species to another. Here is an example that relates not only to the hippocampus but also to a profound issue in the evolution of mind: 'trace conditioning' of the eyelid reflex (*classical conditioning) is sensitive to hippocampal damage and involves *conscious awareness in normal human individuals (Clark and Squire 1998). However, this by itself is insufficient to propose trace conditioning as a cross-species assay for awareness, because other potential explanations (*Ockham's razor) must first be scrutinized, such as a failure to hold information off-line irrespective of awareness.

Another caveat that should be considered is that occasionally, an assay becomes a prison to imagination. This problem runs in two versions: individual and generic. Some individuals flirt with a single method, even a single assay, throughout their career, from their Ph.D. thesis on. Being inflicted with some unique version of separation anxiety, they refuse to give up a procedure that has worked for them, and entrust their future in the hands of the past. A more serious problem arises when an entire subdiscipline falls into the procedural drain. For example, in its first few years, the newly emerging discipline of mammalian neurogenetics has followed as a routine a very limited number of standard versions of the otherwise very useful water maze assay. This was also occasionally accompanied by the simplistic interpretation of performance in the maze, probably resulting in neglect of some intriguing effects of mutations on behaviour (on some of the complexities involved, see Bannerman *et al.* 1995; Wolfer *et al.* 1998).

It is likely that in due time, memory research will generate memory-specific assays based on direct observation of experience-dependent alterations in *internal representations of the nervous system (*map, *functional neuroimaging; example in *honeybee).

Selected associations: Delay task, Maze, Method, Paradigm

Associative learning

1. The formation of new mental links among events.[1]

2. Learning that depends on the parameters of more than a single *stimulus.

The notion of 'association' is central to both the philosophical and the experimental study of the mind. In philosophy it can be traced back to Aristotle, who proposed that similarity, contrast, and contiguity of images subserve recollection (*On memory*; Sorabji 1972). 'Associationism', the philosophical doctrine that the mind learns and construes the world bottom-up by associating elementary events, has emerged with British empiricism in the seventeenth century (Warren 1921). Hobbes (1651) talks about 'the train of thoughts' and of 'compounded imagination … as when from the sight of a man at one time, and of a horse at another, we conceive in our mind a Centaure'. It was, however, Locke (1690) who first used the phrase 'association of

ideas', as the title of a chapter in *Essay concerning human understanding*.

When psychology became an independent empirical discipline towards the end of the nineteenth century, associationism was part of its conceptual heritage. Ebbingahus (1885) was influenced by it when he designed the first quantitative *recall experiments, involving perceptual 'atoms' and their associations. Similarly, Wundt (1896), the founder of the first laboratory of experimental psychology, advocated the study of elementary mental elements and their association in learning, recollection, and thought (Boring 1950a). Over the years the integration of associationism into psychology has also been accompanied by the development of theories[2] that kept the centrality of associations yet disposed of the assumption that the mind works solely bottom-up from simple ideas and 'psychic atoms' (e.g. James 1890; Freud 1901; Hebb 1949; Tversky 1977).

Associations play a part in all the faculties of the mind: learning (the formation of new associations, definition 1); recollection (the use of associations as *cues, *priming, *retrieval); and thought (which involves both the generation of new *internal representations, definition 1, and recollection of old ones).[3] Here we refer to one aspect only, that of *learning. A popular *taxonomy of learning is based on a dichotomy between 'associative' and 'nonassociative' learning. In contrast with associative learning (definition 2), in nonassociative learning, i.e. *habituation and *sensitization, learning is assumed to depend solely on the parameters of the unconditioned stimulus. Whether in *real-life this is indeed the case, is questionable. Even habituation and sensitization involve associations not only with the history of the subject and its interaction with the stimulus, but also with the *context (Hall and Honey 1989; Rankin 2000). Incidental learning and *insight are occasionally depicted as nonassociative as well, but again, this is a great simplification, as in both cases associations are formed in the mind. Incidental learning involves associations between an input and saliency or motivation. Insight is expected to involve sequential implicit associations of internal representations and their *binding. All in all, therefore, it is possible to conclude that associations of some kind or another are universal, and instrumental in learning in even the simplest organisms and tasks.

The study of associative learning has gained tremendously from the use of animal behaviour *paradigms. At the beginning of the twentieth century two major types of paradigms emerged, which permitted for the first time the investigation of elementary forms of associative learning in laboratory animals, and hence a more *reductive and mechanistic analysis of associations at multiple *levels of analysis. One paradigm was *classical conditioning, associated mainly with Pavlov (1927) and his school. The other was *instrumental conditioning or operant conditioning, associated mainly with Thorndike (1911) and later Skinner (1938) and their schools (*behaviourism). In both types of paradigms, the *subject learns relations among events (definition 1). In classical conditioning these relations are among stimuli, whereas in instrumental conditioning, these relations are among actions and their consequences.

The availability of *controllable protocols of associative learning in animals has provided a fertile ground for the development and test of multiple types of laws and theories of associative learning. These theories differ in the identification of the associated variable and of the principles of association. Main types of associated variables considered in these theories are stimulus–stimulus (S–S), stimulus–response (S–R), response–response (R–R), and response–*reinforcer (e.g. see *instrumental conditioning). Stimulus in these theories is commonly an external, sensory stimulus. Note, however, that in definition 2, 'stimulus' is more general and refers to any event that triggers a response in the brain, whether of an external or an internal source, hence it includes also the feedback of motor response. Further, in reality, those are of course not the stimuli themselves that are associated, but rather their on-line *percepts or off-line stored representations. Principles of associations that are considered in theories of associative learning are the frequency of occurrence of the events, their co-occurrence in time and space (contiguity), the probability of linkage (contingency), and the effect or reinforcement (Dickinson 1980; Bower and Hilgard 1981; Mackintosh 1983).

At least in one basic assumption the original British associationism clearly went wrong. This is the depiction of our mental life as dependent only on postnatal associations. Many associations in our brain have innate predispositions. Some authors would even go further to propose that all the associations in our brain are predisposed, and therefore all learning is 'prepared' to some degree or another. This could be due to the existence of certain neural pathways but not others. The generation over time of endogenous pre-representations, which are partially independent of external-world experience but selected by it (Heidmann *et al.* 1984), could also be constrained by *a priori patterns of connectivity in the brain. An example of a simple type of prepared learning is provided by the form of classical conditioning called α conditioning, in which the modified response is pre-existent. Other examples of prepared associations

are *imprinting and *conditioned taste aversion. Whether learning is 'prepared' or not should be taken into account in the search for the cellular and molecular algorithms and mechanisms of learning. For example, presynaptic facilitation of active synapses in the circuit that subserves behaviour (**Aplysia*) fits to subserve prepared learning, whereas the activation of silent synapses or the growth of new synapses fit to subserve *de novo* associations as well.

Selected associations: Classical conditioning, Instrumental conditioning, Priming, Taxonomy

[1]This definition also fits certain artificial systems, such as smart robots, if 'mental' is construed *metaphorically.

[2]As noted in *algorithm, these are not genuine theories in the mathematical sense of the term, but rather conceptual generalizations. The same is true for 'laws' below.

[3]For the role of associations in completing memories from partial input in artificial neural networks *models, see Hopfield (1982), Amit (1989), and Mehrota *et al.* (1997).

Attention

1. **The focusing on part of one's own sensory or cognitive space.**
2. **The selection by the brain of *percepts or longer-lasting *internal representations for *conscious processing and action.**
3. **The selection by the brain of percepts or longer-lasting representations to control ongoing behaviour.**
4. **The alert state required for the above.**

So prominent is the position of attention in the scientific discourse on behaviour, that Titchner (1908) regarded it as 'the nerve of the whole psychological system', and added that 'as men judge of it, so shall they be judged before the general tribunal of psychology'. James (1890) was convinced that 'everybody knows what attention is' and described it as '… the taking possession by the mind in clear and vivid form of one of what seem several simultaneous objects or trains of thought'. James was right in stating that intuitively we know what attention is, but, probably because the concept is so inclusive, a consensus on its definition is not easy to attain.

Not always was attention at the focus of attention of psychology. *Behaviourism intentionally ignored postulated inner faculties of the mind, including attention. The interest was renewed only after the Second World War, with the application of information processing theory, originally developed for warfare purposes, to the cognitive sciences (Broadbent 1958). A large body of work on attention has been accumulated since then, both in psychology and neurobiology. It ranges from investigation of the orienting reflex (*sensitization) to auditory and visual perception. A substantial part of what we currently know on attention stems from the analysis of vision in primates, at *levels ranging from behaviour via *functional neuroimaging and neuroanatomy to single cell activity (Posner and Petersen 1990; Desimone and Duncan 1995; Egeth and Yantis 1997; Kanwisher and Wojciulik 2000).

'Attention' refers to multiple mental states and activities, involving vigilance, orientation, and selection of information. The spectrum of activities thus ranges from the distributed to the selective and to the focused in time and space. These activities engage to various degrees on-line information (percepts of sensory attributes, location and timing) as well as off-line information (i.e. lasting internal representations). Similarly, attention could be *stimulus-driven (a bottom-up process) or task-driven (a top-down process). The latter dichotomy is illustrated in vision. Here selective attention was explained in terms of two consecutive, partially overlapping processes. The first is stimulus-driven, automatic, instantaneous and transient. The second is task-driven, slower, sustained and requires cognitive effort (Sperling and Weichselgratner 1987). Early stimulus-driven processing is frequently referred to as 'preattentive' (Neisser 1967), because it involves parallel processing of primitive features over the sensory space in the apparent absence of mental-resource limitation (Julesz 1981; Treisman 1985). Indeed, central to the notion of attention is resource-limited 'selection' (Norman and Bobrow 1975), which is detected at multiple points between post-receptor input and response (Desimone and Duncan 1995). Hence, lack of resource competition is taken by some authors to indicate lack of 'real' attention. More recent findings suggest, however, that even 'preattentive' vision is constrained by mental resources (Joseph *et al.* 1997).

A common connotation of attention is *conscious awareness (definition 2). Does this mean that attentive nonhuman species can be consciously aware of their dids, and if so, which species? Definitions 1 and 3 above fit situations in which conscious awareness cannot be proven or even assumed. Another definition, suggested by Hebb, also does not specify consciousness: 'central facilitation of the activation of one assembly by the

previous one' (Hebb 1949); this view of attention depends, however, on the validity of the notion of *cell assembly. As far as the relationship of attention to conscious awareness is concerned, it is noteworthy that on the one hand, even humans may not be aware of activity in a cortical area assumed to be involved in some attentional tasks (Crick and Koch 1995); on the other hand, some degree of conscious awareness is expected to exist in other species as well (example in *classical conditioning). It is therefore useful to regard attention as involving a spectrum of awareness. Attention has been proposed to be the *binding agent of consciousness, and it is tempting to speculate that it has been a driving force in the emergence of consciousness. Seen that way, one could not escape the humble conclusion that the most precious niches of our inner world owe their existence to the emergence in evolution of the primitive, elementary orienting reflex.

Developments in two *methodologies have contributed much to the contemporary research on brain mechanisms of attention. One is cellular physiology, used in the *monkey, the other is functional neuroimaging, used in research on human *subjects (Desimone and Duncan 1995; Kawashima *et al.* 1995; Kastner *et al.* 1998; Reynolds *et al.* 1999; Kanwisher and Wojciulik 2000). The combination of both methodologies has led to the identification of brain circuits and cellular processes that are engaged in attention either correlatively or casually (*criterion).

At the system level, research on visual attention shows that areas in the frontoparietal, inferotemporal and occipital *cortex are involved. Among the visual processing areas, high-order cortex is particularly engaged, but there is also evidence for attentional activity already at the primary visual cortex. Attending a stimulus modulates the activity in cortex, even when the subject only expects to attend the stimulus before stimulus onset[1] (Chawla *et al.* 1999). This is taken to reflect the task-driven, top-down attentional facilitation of the processing in the area that expects the signal. There is also evidence for hemispheric lateralization, with a right hemispheric bias for tasks involving attention to locations in space and left hemispheric bias for tasks involving attention to timing (Coull and Nobre 1998). As to the frontal cortex, it is considered to subserve a 'supervisory attentional system' or 'central executive system', which co-ordinates and prioritizes attention across sensory and internal modalities (Shallice 1988; Baddeley 1993). This is the same cortex involved in *working memory. This should not be surprising, since clearly, attention and working memory are complementary and closely related (James 1890; Cowan 1988; Baddeley 1993). Attention identifies where the action is (a popular *metaphor likens it to a searchlight, Crick 1984*a*); working memory then immediately takes note of that action for further use. By so doing, it not only permits an instantaneous *plastic response, but also prevents superfluous exploitation of attentional resources. Whereas some of the automaticity in stimulus-driven attention is innate (*a priori), it is clear that the system has to be capable to quickly compare stimuli with use-dependent internal representations in order to decide whether focused attention and further processing and action are warranted. This interplay of attention and memory takes place within a fraction of a second of perception. Working memory is therefore also 'working attention' (Baddeley 1993).

At the cellular level, attention was found to increase the magnitude of the response of neurons in higher-order visual cortex to the attended stimulus in the receptive field;[2] when multiple stimuli are within the receptive field, the activity is larger when attention is directed at the target stimulus (Moran and Desimone 1985; Reynolds *et al.* 1999). This gain and gating control could involve multiple circuit and system mechanisms, including the action of diffused neuromodulatory systems (*neurotransmitter). The function of these neuromodulatory systems in learning is assumed to involve regulation of gain and gating control as well; hence at the *synaptic level, certain molecular mechanisms of learning and attention merge.

A variety of pathologies impair attention. Among these are parietal and frontal lesions (Shallice 1993), schizophrenia (Andreasen *et al.* 1994), and attention-deficit/hyperactivity disorder, one manifestation of which is learning difficulties (Shaywitz *et al.* 1997). It has been suggested that attention and memory are also co-impaired in chronic fatigue syndrome, and the hypothetical 'central executive' was implicated (Joyce *et al.* 1996). In *real-life, multiple methods could be used to enhance attention, and, good news, some of these methods are clearly devoid of any side effect: a comparison of memory for humorous and non-humorous versions of sentences shows that the humorous ones are remembered better, probably because they are associated with increased attention (Schmidt 1994).

Selected associations: Binding, Homunculus, Metaphor, Percept, Working memory

[1]Expecting to attend is actually an 'attentional set'; for more on what is meant by 'set', see *learning set.

[2]A receptive field is that sector of the sensory space that could be sensed by the neuron.

Behaviourism

1. **The conceptual framework and the school of psychology that consider only overt behaviour as the subject matter of scientific psychology.**
2. **The philosophical stand that considers propositions about mental states identical to propositions about behavioural dispositions.**

The tenet of behaviourism is that behaviour rather than mind or brain is the subject matter of psychology, and that only publicly observed behaviour can be used as psychological datum. Although its roots can be traced to earlier materialistic philosophy and physiology, the formal emergence of behaviourism in psychology is associated with a manifesto entitled 'Psychology as the behaviorist views it' (Watson 1913):

> Psychology as the behaviorist views it is a purely objective experimental branch of natural science. Its theoretical goal is the prediction and control of behaviour. Introspection forms no essential part of its methods, nor is the scientific value of its data dependent upon the readiness with which they lend themselves to interpretation in terms of consciousness. The behaviorist, in his efforts to get a unitary scheme of animal response, recognizes no dividing line between man and brute.

Several points deserve special attention in Watson's manifesto. First, the rejection of introspection as a valid scientific method, opposing a major trend in psychology at the turn of the twentieth century (Boring 1950; Boakes 1984). Second, the rejection of *consciousness as the subject matter of psychology, again, in contrast to contemporary trends (*ibid.*). Third, the emphasis on the phylogenetic continuity, drawing from Darwinism and legitimizing animal psychology as an approach to the study of human behaviour (Boakes 1984). And fourth, aiming at control of behaviour. The latter objective is clearly not a necessary element of behaviourism, but did recur in the history of the field, occasionally endowing it with Orwellian connotations. The pragmatic attitude (Watson ended up in commercial advertising) culminated on the one hand in rather outrageous experimentation on *fear conditioning of human babies (Watson and Rayner 1920), and on the other in attempts to convince pigeons to guide missiles across enemy lines (Skinner 1960). In a more practical endeavour, it also set foundations for behavioural psychotherapy (Wolpe 1963).

Despite recurrent premature elegies, behaviourism retained its vigour over many years. Like other influential concepts, the original notions mutated. Several *taxonomies are noteworthy. One of these classifies behaviourism by period or school. 'Classical behaviourism' is Watson's. It is also dubbed 'molecular', because it treats behaviour in terms of individual 'atoms' of *stimuli, responses, and single stage stimulus–response operations. 'Neobehaviourism', itself a mixed bag, is associated mainly with Tolman (1932), Skinner (1938), and Hull (1943). It treats behaviour in molar terms of classes and types, and its variants incorporate not only stimuli, responses, and *reinforcers (i.e. operations performed on the organism), but also mediating variables that are not directly observable but thought to be necessary for explaining behaviour (see *algorithms). The Skinnerian version of behaviourism (Skinner 1938) is called 'radical behaviourism', although the same term was initially used to denote classical behaviourism (Calkins 1921). It intentionally ignores mind and brain processes (in his later writings Skinner said that brain sciences are indeed relevant, but not useful in analysing behaviour; Skinner 1988). Radical behaviourism advocates a world view in which behaviour is explained in terms of responses to stimuli and modification of probability of responses by contingencies with reinforcements. It disposes of mental causes; the unobservable 'mind' is replaced with mechanistic responses of various complexities, selected either in the species' evolution (*a priori), or by the reinforcement history of the individual *subject. The pinnacle of Skinnerian behaviourism was the attempt to explain human language (Skinner 1957), an attempt ardently rebutted by linguists and cognitive psychologists (Chomsky 1959).

Another taxonomy distinguishes 'methodological' from 'philosophical' behaviourism (on either one or both, see Carnap 1933; Ryle 1949; Zuriff 1986; Collins 1987; Todd and Morris 1995). Methodological behaviourism advocates the aforementioned principle that scientific understanding of the mind has to rest entirely on publicly observable facts, yet without necessarily taking a stand on inner mental realities (definition 1 above). In contrast, philosophical behaviourism does make statements about mental realities (definition 2), which comes in at least two versions: 'metaphysical' and 'logical'. Metaphysical behaviourism makes life easy by denying mental phenomena, period. Logical behaviourism considers propositions about mental states identical to propositions about behavioural dispositions. It can therefore be said to *reduce mental into behavioural acts.

Over the years, behaviourism has experienced fierce attacks from biological and cognitive psychology, linguistics, and philosophy (for arguments related to the insufficiency of behaviourism to account for learning, see Dickinson 1980). As noted above, behaviourism excluded itself from the biological arena in which much of the excitement of modern memory research takes place. Nevertheless, even with the recent developments in the neurosciences, behaviourism is still highly relevant to basic concepts addressed in this book. For example, the mere definition of *memory raises the issue of the relevance of observable facts to inferred processes. Behaviouristic definitions of learning and memory cannot guide neurobiological research because they are not expressed in biologish. But similarly, data on *ion channels and *synapses cannot advance memory research unless they are expressed in a behaviourally relevant language. Skinner (1988) pointed out that 'Sherrington never saw the action of the synapse about which he spoke so confidently'.[1] We do see it now. An aim of modern neuroscience is to observe neuronal function in the context of circuits and neuronal populations (*cell assembly) that encode *internal representations and guide behaviour. The *level of internal representations, which the classical and radical behaviourist tabooed, is hence expected to bridge the organismic and the molecular approaches to memory. We distanced ourselves long ago from the hegemony of introspection that the fathers of behaviourism so much distrusted, but we are still striving to reach the stage in which brain activity will provide accountable, reliable, and objective measures of behaviour.

Selected associations: Culture, Instrumental conditioning, Paradigm, Performance

[1]On Sherrington, see under *synapse.

Bias

1. A preference or inclination that impairs impartial judgement.

2. The favouring of some outcomes over others as a result of systematic errors in procedures or interpretations.

Frances Bacon, trusting that 'the subtlety of nature is greater many times over than the subtlety of the senses and understanding' (Bacon 1620), distinguished four classes of 'idols' (illusions) that beset the human mind: Idols of the 'Tribe' (inherent in the *a priori limited capacity of the species' senses and mind), of the 'Cave' (resulting from the individual's education and experience), of the 'Market-Place' (originating in social influence and public opinion), and of the 'Theatre' (stemming from dogmas and illusory knowledge). The analysis of error and bias in science has since became richer and more sophisticated, but the basic illusions still haunt us: those that stem from the senses, faulty logic, acquired prejudices, and suffocating *paradigms. Science has learned to cope with the shortcomings of the senses, yet finds it rather difficult to struggle with other faults of human nature, be them conscious or not.

Bias could be explicit (definition 1) or implicit (definitions 1 and 2). But even if explicit, it should definitely be distinguished from explicit distortion, which falsifies the data. The latter deplorable disease will not be discussed here further. At the other end of the spectrum stand the 'idols of the tribe', the elementary sensory and cognitive illusions that bias reality and usually transcend culture, education, and profession (Gregory 1966; Kahneman and Tversky 1982); they will not be referred to here either.

In the context of the present discussion, it is methodologically useful to distinguish four major domains in which bias could emerge: The behaviour of the experimental *subject, that of the experimenter, the interaction between the subject and the experimenter, and the scientific community that judges the research project. A notable source of potential *perceptual, *attentional, mnemonic, and judgement bias in the subject, is the emotional state (Power and Dalgleish 1997). For example, depression imposes a bias toward recalling unpleasant rather then pleasant memories (Clark and Teasdale 1982; see also 'mood congruency' under *state-dependent learning). In some situations, interactions unknown to the experimenter among individual subjects in a shared experimental situation, could lead to biased response by the subjects and *artefacts on the side of the experimenter (e.g. Heyes *et al.* 1994; *observational learning). In addition, multiple sources of bias stem from implicit interactions of the subject with the *context and the experimenter. In many behavioural experiments, the subject is actively involved more than the experimenter is inclined to admit. The subject pays attention to the experimental demands, could try to extract *cues about the objective of the test, reacts to involuntary signs emitted by the experimenter (*Clever Hans), and sometimes attempts to comply with a perceived goal (Pierce 1908). The cues that convey an experimental 'hypothesis' to the subject

and hence influence the subject's behaviour are termed 'demand characteristics' (Orne 1962). Their influence on the behavioural outcome of an experiment were mostly studied in humans, but they clearly exist in experiments involving other species as well. Demand characteristics may lead to biased responses by the subjects and to potential artefacts on the side of the experimenter. And finally, the experimenter is itself a potential source of bias (Rosenthal and Rubin 1978; Martin and Bateson 1993). An almost trivial source is self-deception, motivated by a wish to obtain certain results but not others (a potential negative spin-off of *scoopophobia). In such situations minor acts of sampling bias and even data selection throughout the experiment could accumulate to a significant impairment in the overall outcome.

Proper *controls in the experimental design are a must if one wishes to minimize bias due to the subject, experimenter, or experimenter–subject interactions. For example, the potential for some facets of bias could be reduced by strictly following a 'blind' design, in which the person making the measurements does not know the treatment each subject has received until after the experiment is over. In human experiments (such as those that test the effect of drugs on behaviour), a 'double-blind' design should be followed, in which the subject as well does not know the treatment. Furthermore, experimenters must be well aware of their own behaviour. For example, the location and the bodily gestures of the experimenter could markedly bias the behaviour of a *rat or *mouse in a *maze. The design and execution of reliable learning and memory experiments is a complex mixture of science and art, and at least the science part (Martin and Bateson 1993; Kerlinger and Lee 2000) should be mastered before the first experiment is trusted.

But the ordeal of overcoming bias in the experimental design and in its execution is not over even when the manuscript is finally ready for publication. The idols of the market-place and of the theatre could still pose substantial obstacles. The attitude of referees and editors is sometimes biased by *zeitgeist, by a prevalent conceptual paradigm, or, even worse, by the fame of the senior author or the institution in which the work had been done. The refusal over years to accept papers on *conditioned taste aversion, because it had seemed to defy some ideas about what conditioning should be (Garcia 1981; *classical conditioning), provides but one example of referees and editors being biased by a conceptual paradigm. In other cases, the wish of referees and editors to appear politically correct in their scientific milieu or in society at large may also introduce bias into the scientific literature.

Selected associations: Control, Culture, Paradigm, Subject

Binding

1. The phenomenon or process in which elements in space and time cohere into a perceived whole.

2. The phenomenon or process in which perceptual features integrate into a coherent sensory *precept.

3. The phenomenon or process in which representational elements fuse into a coherent *internal representation.

There are several facets of 'binding' (definition 1) that excite philosophers. A classical aspect has to do with the persistence of the identity of things whose constituents turn-over with time, such as the self (see the 'Ship of Theseus' problem in *persistence). Neuroscience and the philosophy of mind presently focus on a distinct type of the 'binding problem', which refers to the ability of the brain to bind, within a fraction of a second, the features of a complex *stimulus into a coherent, meaningful percept (definition 2). Interest in this type of problem has a long history (e.g. Hume 1739). Neuroscience has dragged it into the laboratory, although for many scientists it still retains an excessively 'soft' connotation (also see below). Consider vision: different types and combinations of visual attributes are processed in the brain in multiple streams (Knierim and Van Essen 1992). How do they recombine to yield a coherent visual percept (Treisman 1993; Singer and Gray 1995; Shadlen and Movshon 1999)? This is the 'Humpty Dumpty' problem: 'Humpty Dumpty sat on a wall/ Humpty Dumpty had a great fall/All the King's horses and all the King's men/ Couldn't put Humpty together again' (Carroll 1872). In the brain Humpty is put together again. Or at least so we sense. How sad is it not to: 'On an incredibly clear day/… I saw …/That Great Mystery the false poets speak of…/That there are hills, valleys and plains/That there are trees, flowers and grass/There are rivers and stones/But there is no whole to which all this belongs/ That a true and real ensemble/ Is a disease of our own ideas.' (Pessoa 1914).[1]

Binding is related to the coherency of all kinds of internal representations (definition 3), not necessarily in the context of sensory perception. Hence it

surfaces, either implicitly or explicitly, in discussions of *memory (Squire *et al.* 1984; Teyler and DiScenna 1986; Damasio 1989; Hommel 1998; Dudai and Morris 2000). These discussions usually refer to *declarative memory, but sometimes generalize to simple stimulus–response representations. Clearly, although the mainstream interest in the 'binding problem' is still in the context of perception, whatever will be gained there will contribute to the understanding of memory as well.

The binding problem binds several subproblems. Here is a selection:

1. *Parsing:* How are the relevant elements selected among other elements in the perceptual or mental space (Treisman 1999)? And how much of this selection is constrained by *a priori rules?
2. *Encoding:* How is the binding marked, maintained, and read by other systems in the brain (*ibid.*; *cell assembly)?
3. *Mapping:* How are the elements, once bound, kept in the correct structured relations (*ibid.*; *map)?
4. *Flexibility:* How are the bound elements reused in binding without lingering interference of the previous binding(s)?

Each of these questions could be tackled at multiple *levels, from that of the computational theory, via the *algorithms that implement the computations, down to the biological hardware that implements the algorithms. Discussion of binding in cellular neurobiology is still rather uncommon. The main focus is on the higher levels of neuronal circuits, brain systems, and cognition. At these levels, it is methodologically convenient to distinguish two types of approaches: top-down or cognitive, and bottom-up or neurobiological.[2] The *classic top-down approach is that of the Gestalt School (*Gestalt*, from German for '*shape*'; Koffka 1935; Hochberg 1998; *insight). This school of psychology, founded in Germany in the early twentieth century, has promoted the view that the nature of perceptual parts is determined by the whole, and that enquiry into the mind should consider global organization and proceed top-down. Unfortunately not much top-down analysis of the brain was possible during the formative years of the Gestalt. In more recent cognitive psychology, an influential *model is that of 'feature integration' (Treisman and Gelade 1980; Treisman 1993). This model considers *attention as the binding agent. It proposes that simple perceptual features are registered in parallel across the visual field, in a number of specialized subsystems. Focused attention scans serially, within milliseconds, through a 'master-map' of locations, accessing the features present there at that point in time. The features are integrated, or 'glued', by the attentional 'beam' (*metaphor; for critical discussions of 'feature integration', see M. Green 1991; Van der Heijden 1995; Treisman 1995).

The neurobiological approach attempts to identify computations and algorithms relevant to binding in the brain and their physiological implementation. It focuses on the *cortex and on thalamocortical interconnections; in discussions of the role of binding in memory, attention is also devoted to the *hippocampal formation and to its role in coherency of internal representations. Two major types of solutions come up in neurobiological models of binding. The first type of solution is that binding is based on a *place code* (*map), and is performed by hierarchical combination of coding units, which converge anatomically on a master location (*homunculus; Barlow 1972; also discussions in Singer and Gray 1995; Grossberg *et al.* 1997; Bartels and Zeki 1998). The second type of solution proposes that binding is based on a *temporal code* (Eckhorn *et al.* 1988; Hardcastle 1994; von der Malsburg 1995; Engel *et al.* 1997). The basic idea in this case is that feature-detecting neurons are bound into coherent representations of objects if they fire in synchrony. Neurons in the cortex have been indeed observed to engage in recurrent bursts at frequencies of 30–70 Hz, and this has specifically been proposed as a candidate mechanism of binding. It also fits psychophysical data, which suggest 20–30 ms as the time scale of a 'cognitive beat' (*capacity, *percept). At this stage, the temporal synchrony hypothesis is still mostly phenomenological. It is not yet clear whether the oscillations represent a causal mechanism, a phenomenon, or an epiphenomenon (*criterion). To understand what's going on, one would wish to identify the semantics of the representational code(s), the source of the oscillations (i.e. intrinsic, emergent ensemble properties, top-down induction or executive control), and the hardware components (e.g. *coincidence detector).

So is 'binding' as defined above a problem, or a pseudoproblem? The same question applies to other *enigmas of the brain. What distinguishes 'binding' from some other unresolved brain processes and mechanisms, and occasionally endows it with a mystic flavour, is probably its association with major philosophical aspects (or some would say 'spin-offs') of the neurosciences. These include the mind–body problem and *consciousness (e.g. Crick and Koch 1990). Many scientists hesitate to touch these issues, others do it rather enthusiastically. Crick (1994) remarks on the 'binding problem' that 'it is not completely certain that this is a real problem or the brain gets around it by some

unknown trick'. Sure the brain does its trick, and the problem is hence only ours to solve. 'That wonder is the effect of ignorance has been often observed' (Johnson 1751). It is therefore likely that with time, the 'binding' will stay but the 'problem' dissipate.

Selected associations: Algorithm, Attention, Cell assembly, Coincidence detection, Percept

[1]Pessoa's lines seem to echo a neurological disorder, Balint's syndrome, in which the ability to perceive the visual field as a whole is disturbed due to bilateral damage to the occipitoparietal region (Halligan and Marshall 1996).

[2]Bottom-up analysis of perceptual binding commonly attempts to account for cognitive phenomena by circuit and multicircuit properties. In the process, it could still employ top-down analysis of synaptic properties.

Birdsong

Complex, stereotyped vocalizations, accompanied by characteristic body postures, produced predominantly by mature male birds during the breeding season.

Male birds sing to selected audiences. The male is a landlord and potential warrior, notifying other males that it is ready to defend its territory. It is also a charming troubadour attempting to convince females that it is the best in town. The song occupies such a cardinal role in the male's life that it may even dream about it (Dave and Margoliash 2000). Whereas we humans could enjoy the song repertoire regardless of gender, the male and the female of songbirds are probably each tuned to understand only that part of the song that speaks to their heart (Williams and Nottebohm 1985). The *plasticity of birdsong has been well known to bird fanciers in the Orient since ancient times, and expert manipulations of song were exploited for aesthetic and commercial purposes (Konishi 1985). This neuronal and behavioural plasticity has also long attracted scientists' attention (Darwin 1871; Mertfessel 1935; Koehler 1951; Thorpe 1954). In addition to being a beautiful system to investigate ethology and learning, the study of birdsong taps into several central issues in brain research. These include the role of genetic constraints on learning ('prepared learning', see *a priori, *imprinting); the interplay of *development and learning; the contribution of 'instruction' and 'selection' processes in learning (see *a priori, *stimulus); and the role of neurogenesis in the adult brain.

A song is a series of sounds with silent intervals between them. It is different from a 'call' that is a simple, brief vocalization uttered by both species in all seasons in response to particular stimuli such as a predator. Calls are not unique to birds. Birdsongs are. The most elementary sound in a song is a note, lasting 10–100 ms. Notes form syllables, syllables phrases, and phrases songs. Songs are commonly 1–5 s in duration. Different songs form a repertoire. The size of the repertoire ranges from one to many hundreds songs, depending on the species. Repertoires of geographically distinct populations of the same species often differ, and are termed 'dialects' (Baker and Cunningham 1985). Some species perform all or most of their repertoire in cycles that takes many minutes to complete (Marler 1984). Terms such as 'dialect' should not lure us to regard birdsong as an analogue of human language, as there is much more to language than structured stereotyped vocalization. But still, the song repertoire provides the bird with a complex expressive and communicative system, which may require special strategies to ensure prompt *retrieval and correct response (e.g. Todt and Hultsch 1998).

The ontogenesis of song involves discrete stages. Take the wild chaffinch as an example (Nottebohm 1970). In the spring, immediately after hatching, chaffinches begin to emit various food-begging calls. Within a few weeks, the male starts to emit a loose, rambling aggregation of low volume notes of varying complexity. This vocal pattern is called 'subsong'. The subsong keeps changing, and discrete passages, resembling the adult song, gradually emerge. These passages are called 'plastic song'. During the breeding season the subsong vanishes and the plastic song crystallizes into the full adult song. The singing posture typical of the adult also matures. The final crystallization takes place before the end of the winter.

Although there are remarkable species differences in song development, data from experiments involving sensory and social isolation (e.g. Marler and Tamura 1964; Konishi 1965) generalize to portray the following *model of song ontogeny: the bird is born with a song motor-control system that needs input in order to generate a normal song. This input is provided in two stages, 'sensory' and 'sensorimotor', which may partially overlap, depending on the species. First comes the 'sensory stage', during which the bird listens to a tutor. There is a genetically determined predisposition to prefer a conspecific tutor. Thus, even if we raise a chaffinch in Pavarotti's house, the chances that it will learn to sing *La Bohème* are very slim indeed. In the

sensory stage, elements of the tutor's song are confined to memory. In the 'sensorimotor' stage, which corresponds to the subsong, plastic song, and crystallization, the bird must listen to itself to match its vocal output with its innate template as well as with the memorized template of the tutor's song. The entire process combines elements of instruction (by the tutor) and selection (among endogenous innately constrained song templates; see Marler 1997). In the absence of a tutor, only the innate information is used. Some species can generate species-specific song solely on the basis of an innate template, in the absence of tutors and auditory feedback. Species differ also in the stability of song. In some 'age-limited learners', such as the zebrafinch and white-crowned sparrow, learning is limited to the first year of life, and crystallized song is maintained throughout adulthood. In 'open-ended learners', such as the canary, new songs are added in adulthood. But even adult 'age-limited learners' retain a significant amount of plasticity, and use auditory feedback in adulthood to maintain the stability of song structure (Leonardo and Konishi 1999).

One of the advantages of birdsong as an experimental *system is the well defined and quantifiable behavioural output that provides a convenient and faithful *assay to determine whether learning has occurred. Moreover, song is generated by a single organ, the syrinx. This facilitates the tracing of pathways from central motor centres and ultimately identification of brain circuits that subserve *acquisition and execution of the motor programme. Over the years, in series of elegant studies combining anatomical and cellular *methods, a picture has been generated that depicts the major elements of the central song system as composed of two major forebrain pathways (Figure 6). The posteriomedial pathway is traditionally termed 'the motor pathway', and includes, in ascending order, the nucleus Uva, the nucleus NIf, the higher vocal centre (HVc, originally so abbreviated because it was thought to be 'hyperstriatum ventrale'), and finally nucleus RA, that innervates the tracheosyringeal portion of the hypoglossal nerve nucleus (nXIIts), itself innervating the syrinx. This pathway is fed by auditory input, and is obligatory for both song development and production (Nottebohm *et al.* 1976). Lesions in HVc and RA result in 'silent song': upon noticing a female, the lesioned male adopts a singing position but emits no song, becoming a very sad bird indeed. The HVc and RA are organized hierarchically, with HVc neurons representing syllables and RA neurons representing notes. Uva and NIf may help organize syllables into higher units of song. Some of the sites afferent to Uva

Fig. 6 A schematic representation of the songbird brain, showing the brain centres and pathways that subserve the development, learning, and production of song. The *system is composed of two major forebrain pathways. The posteriomedial ('motor') pathway (*black*) includes the nuclei Uva, NIf, HVc, and RA. The RA innervates the tracheosyringeal portion of the hypoglossal nerve nucleus, which in turn innervates the song organ, the syrinx. The anterior forebrain pathway (*grey*), which is obligatory for song development and learning, connects the HVc to RA via area X, the thalamic nucleus DLM, and nucleus lMAN. Also shown is the auditory area L, which feeds into the HVc. Abbreviations: AM, nucleus ambiguus; DLM, medial portion of the dorsolateral nucleus of the thalamus; HVc, higher vocal centre in the neostriatum; lMAN, lateral portion of the magnocellular nucleus of the anterior neostriatum; Field L, auditory region in the neostriatum; NIf, nucleus interface; RA, robust nucleus of the archistriatum; RAm, nucleus retroambigualis; Uva, nucleus uvaeformis; V, ventricle; X, area X; n XIIts, tracheosyringeal part of the hypoglossal nucleus. (Adapted from Brenowitz et al. 1997.)

may also take part in sensory acquisition during song development (Margoliash 1997).

Another interconnected pathway, the anterior forebrain pathway, is considered essential for song development, learning and recognition. It is not obligatory for the mature song production, but still plays a part in feedback evaluation and adaptivity of singing in the adult bird (*ibid.*; Brainard and Doupe 2000). This pathway indirectly connects HVc to RA via area X, the thalamic nucleus DLM, and the nucleus lMAN. All in all, the song system is distributed over nuclei and circuits, and no single site 'stores' the entire score (*engram, *metaphor). Furthermore, a clear-cut dissociation between central 'sensorimotor' and 'learning centres is probably not honoured by the brain.

An intriguing finding is that many new neurons are born in the brain of the adult bird (Goldman and Nottebohm 1983; Alvarez-Buylla and Kirn 1997). Such neurogenesis is not limited to song nuclei, to males, or to species that sing. However, in songbirds, it is prominent in HVc, and correlates with seasonal variations in

song and sex hormone levels. (Sex hormones play a part in moulding song circuits and behaviour; Bottjer and Johnson 1997.) The role of neurogenesis in song memory, if at all, is not yet clear. In recent years neurogenesis has also been noted in the adult mammalian brain, and, furthermore, reported to be enhanced in learning (Gould *et al.* 1999; *hippocampus; but see concerns in Rakic 2002). Neurogenesis in birds in general and songbirds in particular may therefore reflect a more general process. This is surely a finding that can defeat the popular notion that old brains only fade out.

Selected associations: Development, Engram, Imprinting, Observational learning, Skill

Calcium

A metallic element that comprises about 3% of the earth crust and is essential for many biological processes, including neural *plasticity.

Calcium (*calx*, Latin for lime) fulfils many regulatory, computational, and representational functions in the brain. Furthermore, it is instrumental in translating information across *levels and time domains in the brain (see below). In recent years much has been learned about the ways in which calcium ions (Ca^{2+}) encode and modulate neuronal information, but the picture is far from being comprehensive.

In resting cells, intracellular Ca^{2+} is in the range of 10–100 nanomolar. Upon stimulation it could rise by several orders of magnitude. In many cases the information in the Ca^{2+} signal is encoded as spatiotemporal patterns of change rather than a tonic increase in concentration. Changes in cellular Ca^{2+} are due to influx from the extracellular milieu and release from intracellular stores. Both mechanisms generate elementary all-or-none Ca^{2+} signals, which are brief and localized (Bootman and Berridge 1995). Stimulus-induced combinations of intensity, timing, and location of these primitives of the 'Ca^{2+} language' generate a repertoire of Ca^{2+} codes (Bootman *et al.* 1997). The latter control cellular metabolism, structural dynamics, signal transduction, hormone release, differentiation, and growth (Berridge 1993; Petersen *et al.* 1994; Ghosh and Greenberg 1995; Matthews 1996). The introduction of novel technologies of molecular biology, cellular electrophysiology, and imaging has opened new vistas in the analysis of Ca^{2+} in neurons. Especially noteworthy in the context of plasticity are the studies on the role of Ca^{2+} in mediating and modulating excitability and integrative properties in dendritic compartments (Markram *et al.* 1995; Magee *et al.* 1998); control of *neurotransmitter release (Matthews 1996; Goda and Sudhof 1997); modification of membrane *receptors (Barria *et al.* 1997); and modulation of gene expression (Bito *et al.* 1996; Dolmetsch *et al.* 1998).

The ubiquitousness of Ca^{2+} signalling in the nervous system makes it impractical to mention all its major functions in experience-dependent neuronal modification. These functions are performed at locations ranging from neuronal subcompartments to circuits, and on time-scales ranging from milliseconds to days and more. Ca^{2+} is required for elementary short-lived processes of *synaptic plasticity (Thomson 2001), and for the induction of *long-term potentiation, a popular cellular *model of longer-term neuronal plasticity (Nicoll and Malenka 1995). A few examples will serve to illustrate the role of Ca^{2+} in *acquisition, retention, and consolidation of learned behaviours. In the circuits that subserve *classical conditioning of defensive reflexes in **Aplysia*, Ca^{2+} encodes information about the conditioned stimulus (CS). Furthermore, convergence of the CS and the unconditioned stimulus (US) takes place on a Ca^{2+}/calmodulin-activated adenylyl cyclase (*coincidence detection; *intracellular signal transduction cascade). The optimal activation of the enzyme requires that Ca^{2+} preceded the transmitter, hence mimicking the order dependency of CS–US presentation in classical conditioning (Yovell and Abrams 1992). Another Ca^{2+}-regulated enzyme, the multifunctional Ca^{2+}/calmodulin activated *protein kinase type II (CaMKII; Braun and Schulman 1995; De Koninck and Schulman 1998), was found to be essential in learning (Bach *et al.* 1995), *long-term potentiation (Barria *et al.* 1997), and neuronal development (Wu and Cline 1998). CaMKII is a major component of the postsynaptic density. It phosphorylates and modifies receptors, channels, and cytoskeletal elements. Experience-dependent autophosphorylation of the enzyme complex was proposed as a molecular storage mechanism immune to molecular turnover (Miller and Kennedy 1986). Another family of Ca^{2+} regulated protein kinase, PKC, was also implicated in learning (e.g. Scharenberg *et al.* 1991.) In addition, Ca^{2+} is involved in cellular consolidation: it regulates the activity of *CREB, and hence of the expression of cyclic adenosine monophosphate-response element (CRE)-regulated genes (Bito *et al.* 1996; *immediate early genes).

Why is it that Ca^{2+}, rather than any other ion, plays such a key part in cellular activity in general and in plasticity in particular? Though in essence a teleological question with speculative answers, it does warrant consideration, because it could illuminate interesting properties of Ca^{2+} signalling systems. Possibly the physicochemical parameters of Ca^{2+}, when considered in combination with those of critical Ca^{2+}-binding sites in the cell, had from the early days of evolution fitted the demands of cellular function and plasticity better than those of other ions. The problem with this line of reasoning is that it is of the egg-and-the-hen type: was the cause the abundance of Ca^{2+}, or the availability of the biological binding sites? This inherent issue notwithstanding, one appealing argument in favour of Ca^{2+} at the current stage of evolution is that the affinity of Ca^{2+} for important macromolecules in the cell is strong enough to allow rapid binding but not too strong to prevent rapid dissociation. This is important in cellular signalling in general and in fast plasticity in particular. For example, magnesium binds stronger to phospho-groups (Dawson *et al.* 1986); and monovalent ions are in general much worse in getting bound to biological macromolecules. The problem is highly complex, because, as mentioned above, it is not tonic Ca^{2+}, but rather Ca^{2+} transients, which are most important in signalling. The life-span of these transients may not be sufficient for Ca^{2+} to equilibrate with binding sites in the cell (Markram *et al.* 1998b). Analysis of Ca^{2+} signalling, therefore, requires gigantic calculations of nonequilibrium Ca^{2+} dynamics. For our purpose suffice it to remember that the real-life role of Ca^{2+} in neuronal plasticity must be considered in the context of the simultaneous interaction of this ion with the network of the many Ca^{2+} binding molecules in the neuron.

It is also noteworthy that overall, the actions of Ca^{2+} in the neuron span orders of magnitude in time, space, and complexity (Bootman *et al.* 1997). This endows Ca^{2+} with a unique position to bridge molecular, cellular, and system levels of brain action (Dudai 1997b). The spatiotemporal pattern of Ca^{2+} is therefore a candidate parameter for future equations of the not-yet-available interlevel 'correspondence rules' in brain models and theories (*reduction).

Selected associations: Intracellular signal transduction cascade, Ion channel, Plasticity, Reduction, Stimulus

Capacity

1. The ability of a *system to receive, process, store, represent or transmit items.

2. The measure of this ability.

3. The upper limit of this ability.

Pondering the capacity of our memory carries with it the risk of being enslaved to the common *metaphor of memory as a static storehouse (Roediger 1980). This misconception should be avoided at the outset. Furthermore, in the case of the nervous system, even the definition itself evokes cardinal issues: What is the meaning of 'store' (definition 1)? Are *internal representations stored as such, reactivated, or reconstructed anew each time they are *retrieved?[1] If memory is reconstructed, then the capacity of the system should involve the ability to decompress and recreate information; however, something must eventually be stored as, clearly, the brain does not reconstruct memories from void. And as if all this is not enough, it is likely that different memory systems encode information in different ways, possess different capacities, and exploit the capacity to different extents. Having said all this, it is still of interest to wonder whether in terms of capacity (definition 3), our brain is any match to a notebook computer.

The data are still scarce. The Swiss-German physiologist Haller, who in the eighteenth century performed the first documented experiments on the timing of psychic processes, reached the conclusion that a third of a second is sufficient time for the production of one idea. Hence assuming only eight mentally useful hours per day (!), in 50 years a person has a chance to collect up to 1 577 880 000 traces (Burnham 1889). More recent (yet not necessarily less controversial) estimates of how much information we perceive during an average lifetime, yield the very wide range of $\alpha 10^{13}$–$\alpha 10^{17}$ bits (reviewed in Dudai 1997a; 'bit' is the basic unit in information theory; see *system). In considering the information that becomes available to the brain, we must take into account not only the information that is obtained from the external world, but also that information that is generated endogenously by the brain (*a priori, *internal representation, *stimulus). We do not yet have the bases to estimate the magnitude of contribution of this type of information to the potential representational pool of the brain. *Modelling of artificial 'neuronal' networks of the estimated size of the human brain yields an upper representational capacity of $\alpha 10^{13}$ (Palm 1982) to $\alpha 10^{15}$ bits (Amit 1989). There have been also attempts

to estimate the representational capacity of parts of the brain, such as *cortex (Gochin et al. 1994; Rolls *et al.* 1997). The conclusion was that the available representational capacity is probably more than required to subserve our actual mental and behavioural repertoire.

But how much of this information could be stored in our memory over time? Some agreement exists only on the maximal capacity of short-term, or better, *working memory (*phase). The discussion digresses here from the bits of the formal models to vague, almost impressionistic units. The most popular estimate is that our working memory can hold only seven-plus-minus-two chunks of information at one time. This estimate stems from experiments in psychology (Jacobs 1887; Miller 1956) and from observations in anthropology (Wallace 1961; Berlin 1992).[2] Despite the catchy title of Miller's classic article, seven-plus-minus-two is not a sacred number. There are lower estimates as well (down to only three separate registers; Broadbent 1975). Miller's idea was not to determine a precise value, but rather to point out that the brain is an information processing system of limited capacity, which had evolved to recode information into chunks in order to be able to deal with it efficiently (Baddeley 1994; Shiffrin and Nosofsky 1994). Attempts have been made to estimate the size of a chunk in terms of digits, syllables, words, and patterns (Simon 1974). Some individuals develop a remarkable *skill for chunking, and by combining it with efficient *retrieval from long-term stores, can handle huge amounts of information simultaneously (e.g. more than a 10-fold increase in the normal digit span; Chase and Ericsson 1982).

It has been estimated by Simon (1974), on the basis of the contemporary psychological literature, that 5–10 s are needed to transfer a chunk from short- into long-term stores. When it comes to both the maximal and the actual capacity of the latter, the issue of magnitude becomes even more evasive. In what units should long-term memory be measured? Which 'chunks' should be used to estimate the size of, say, an *episodic scene or a motor skill? Furthermore, how can one compare the capacity of different long-term memory systems? A variety of experimental methods have been deployed, ranging from introspection (Galton 1879), via controlled recalling of personal experience (Wagenaar 1986), to measurement of *real-life capabilities such as picture *recognition, language, or the feats of *mnemonists (Table 1). There are no definite answers, only estimates expressed in *ad-hoc*, somewhat fuzzy units. A conservative estimate is that a normal human long-term memory retains $\alpha 10^5$–10^6 items, where item means a word, a fact, an autobiographical episode—what might intuitively be called a unit of memory, but formally is very unsatisfactory indeed (Dudai 1997).

Table 1 Estimates of the actual capacity of selected human long-term memory stores

Store	Size	Reference
Words in language (mother tongue)	25 000–50 000	Nagy and Anderson (1984)
Pictures recognized	> 10 000	Standing (1973)
Game patterns by a chess master	10 000–100 000	Chase and Simon (1973)
Facts by mnemonists	100 000	Yates (1966)
Core personal episodes	Thousands	Dudai (1997a)
Items in expert databases in orally-reliant societies	500–2000	Levi-Strauss (1966); Berlin (1992)

The capacity of brains and memory systems is no doubt of interest, but it would do no harm to scrutinize the assumptions that underlie this interest. One assumption, which is definitely wrong as a *generalization, is that the bigger, the better. The capacity of memory systems is the outcome of the interplay among multiple drives and elements. These include the functions that this memory system is supposed to accomplish; the mechanistic constraints imposed by the biological machinery; the feasibility of *algorithms; the energy resources that are required to *develop and operate the system; and, finally, the current stage in the evolution of the system. Here is but one concrete example: is it phylogenetically advantageous for the system of *declarative, autobiographical memory to have a large capacity? Not necessarily (see in *false memory).

It would be naive to expect real advances in the estimation of memory capacity before two developments materialize. First, we must decipher the codes of internal representations, in order to be better equipped to estimate the requirements for representational and computational space in the brain. Second, we must gain a much better understanding into the processes and mechanisms of *persistence, *forgetting, relearning in *extinction, and particularly, retrieval of memory. Retrieval that tolerates liberal reconstructions of internal representations, and is heavily dependent on online information, is expected to place different demands on capacity than retrieval that involves faithful reactivation of fine-grained stored information. The issue of capacity is hence intimately associated with some of the most profound *enigmas of memory research.

Selected associations: Episodic memory, Internal representation, Persistence, Working memory

[1]This issue is further discussed in *persistence.

[2]By the way, the working-memory capacity of the chimpanzee is not much less: >5 items, the same as preschool children (Kawai and Matsuzawa 2000).

Cell assembly

A hypothetical concept referring to *phasic sets of coactive neurons that are assumed to encode *internal representations and perform computations over representations.

In 1949, Hebb published *The organization of behavior*, later to become the most influential book in the history of modern neuroscience (*classic). 'In this book', he wrote, 'I have tried … to bridge the gap between neurophysiology and psychology'. In essence, Hebb's monograph was about how the brain *perceives and represents the world. It has yielded important insights into brain function, as well as two major concepts. Typical of Hebb's integrative view of the brain, these concepts related to two *levels: the *synaptic and the *system. At the synaptic level, Hebb coined a postulate of use-dependent synaptic *plasticity (see *algorithm). At the system level, he proposed the existence of neuronal assemblies as vehicles for perception, *attention, *association, memory, and thought. Hebb (1949) envisaged that in the brain

> … stimulation will lead to the slow development of 'cell assembly', a diffuse structure comprising cells in the cortex and diencephalon … capable of acting briefly as a closed system, delivering facilitation to other such systems … A series of such events constitutes a 'phase sequence'—the thought process. Each assembly action may be aroused by a preceding assembly, by a sensory event, or—normally—by both. The central facilitation from one of these activities on the next is the prototype of 'attention'.

Although with time Hebb's synaptic postulate may have gained more popularity (despite being regarded by Hebb himself as less original; Milner 1986), it is the 'cell assembly' that was at the heart of his seminal book. In the past 50 years or so, the concept of 'cell assembly' has remained viable in both experimental and theoretical research on perception, learning, and memory (e.g. Palm 1982; Crick 1984a; Dudai *et al.* 1987; von der Malsburg 1987; Gerstein *et al.* 1989; Singer *et al.* 1990; Nicolelis *et al.* 1997; Sakurai 1998).

The platonic cell assembly has the following attributes: (a) it encodes internal *representations*, in a *spatiotemporal* code; (b) a representation is *distributed* over many units in the set; (c) each unit may be a member of *several* assemblies; (d) the units in the assembly become *coactive*, and hence actualize the assembly and what it represents, in brief time-locked *phases*;[1] and (e) the assembly is *plastic*, meaning that the representations could change over time, either in response to input or by endogenous rearrangements. That the cell assembly uses a distributed, alias ensemble, alias population code means that in big-enough assemblies, no single neuron is essential to any percept or memory; put in other terms, the assembly denies the existence of single-cell *homunculi.

Hebb's assemblies did not emerge out of the blue. As is the case with other great ideas, this one as well stood on the shoulders of giants.[2] The possibility that sensorimotor information is processed by populations of neurons was raised much earlier (Young 1802). Sherrington, the great advocate of the cellular view of brain function, assumed that individual neurons do not have the representational complexity to account for higher properties of the nervous system (Sherrington 1941). Hebb was a student of Lashley, who attempted in vain to localize memory traces to specific brain regions, and reached the conclusion that the *engram is widely distributed (Lashley 1929). At about the same time, de No (1938), himself relying on earlier observations, singled out the role of neuronal loops and recurrent circuits in information processing in the nervous system. This was contrary to contemporary naive switchboard *metaphor, which described the brain in terms of many yet rather simple (sensory) input–(motor) output connectors. Hebb took the aforementioned ideas further. He formulated a comprehensive conceptual framework of brain function in which populations of neurons represent information about the world. As representations (and hence memories) are distributed over many nodes, localized lesions could fail to abolish memory. Furthermore, assemblies according to Hebb are dynamic entities. They form, *develop (first in the immature and later in the mature brain), associate, and disengage. This calls for synaptic plasticity; Hebb's famous synaptic postulate, mentioned above, was his solution to the mechanism of use-dependent modifications in local nodes in the assembly. The first attempts to model