of the shock US. In this paradigm, the amygdala is essential for the acquisition and retention of the CR and this is true regardless of the CS modality (e.g. auditory, visual, olfactory). Lesions of the hippocampus, for example, do not impair fear conditioning using the delay paradigm. However, if the delay paradigm is changed to a trace paradigm by inserting a trace interval of several seconds, lesions to both the amygdala and the hippocampus severely impair the acquisition and retention of the fear CR.

3. Cerebral substrates of eyeblink conditioning

Eyeblink conditioning is the most widely studied form of associative learning in mammals (Thompson 2005). In the basic paradigm, a tone CS is paired with a reflex-eliciting US such as a puff of air to the cornea. Initially the CS does not elicit an eyeblink response. With repeated pairing of the CS and US an association is formed such that that presentation of the CS elicits an eyeblink CR in advance of the US. In general, it takes many more trials to establish a well-formed eyeblink CR than it does the non-specific freezing CR in fear conditioning. Extensive investigation into the neural substrates of eyeblink conditioning using most often rabbits, but also humans, monkeys, and rodents has resulted in perhaps the most complete description of mammalian memory formation to date.

The acquisition and retention of delay eyeblink conditioning requires the cerebellum and associated brainstem structures. These structures are necessary and sufficient for the formation and storage of the CR. No forebrain structures, including the hippocampus, are required. For example, decerebrate rabbits with no remaining forebrain tissue (i.e. after removal of cerebral cortex, basal ganglia, limbic system, thalamus, and hypothalamus) exhibit normal retention of delay eyeblink conditioning. Findings in humans are completely consistent with the animal work. Thus, delay eyeblink conditioning is impaired in patients with cerebellar or brainstem lesions, but intact in amnesic patients with damage that includes the hippocampus (see *brain damage).

However, in eyeblink conditioning, changing delay conditioning to trace conditioning by inserting a trace interval as brief as 500–1000 ms substantially changes the brain substrates and cognitive processes required to support this form of conditioning. For example, successful trace eyeblink conditioning, like delay conditioning, requires the cerebellum. However, trace conditioning differs from delay conditioning in that it also requires the hippocampus and portions of neocortex. Thus, acquisition and retention of trace conditioning are severely disrupted in rabbits and rats when the hippocampus is damaged and trace conditioning is also disrupted by damage to portions of the prefrontal cortex. Again, findings in humans are consistent. In amnesic patients with damage that includes the hippocampus, trace eyeblink conditioning is mildly impaired with a trace interval of 500 ms, and severely impaired with a trace interval of 1000 ms.

4. Awareness and eyeblink conditioning in humans

From a behavioural perspective, work with experimental animals is limited to examining the acquisition, storage, and generation of the CR. However, in humans it is also possible to determine if the participants have additionally developed an awareness that the CS comes before the US and is predictive of the US. Thus, humans have the potential to become aware of this contingency and to develop an expectation of the US following the presentation of the CS. The awareness and expectancy can then be related to the CR in both the delay and trace paradigms (Clark et al. 2002). In both delay and trace conditioning paradigms, individuals sometimes develop awareness regarding the stimulus contingencies and sometimes do not. For the most commonly studied forms of delay conditioning this awareness is superfluous to the acquisition of the CR, presumably because cerebellar and brainstem circuits can support performance. Trace conditioning is fundamentally different. Unlike delay conditioning, trace conditioning is strongly related to the awareness of the CS–US contingency and to the degree to which the US is expected.

It is possible that trace eyeblink conditioning may additionally require the hippocampus and the development of contingency awareness because the trace interval makes it difficult for the cerebellum to associate the CS and the US. In trace conditioning, because the US follows the CS by as much as 1000 ms the cerebellum may not be able to maintain a representation of the CS across the trace interval. If, however, the hippocampus and neocortex have represented the stimulus contingencies, then perhaps processed information concerning the CS can be transmitted to the cerebellum at a time during each trial that is optimal for cerebellar plasticity.

ROBERT E. CLARK

Clark, R. E., Manns, J. R., and Squire, L. R. (2002). 'Classical conditioning, awareness, and brain systems'. *Trends in Cognitive Science,* 6.

Fanselow, M. S. and Poulos, A. M. (2005). 'The neuroscience of mammalian associative learning'. *Annual Review of Psychology,* 56.

Pavlov, I. P. (1927). *Conditioned Reflexes; An Investigation of the Physiological Activity of the Cerebral Cortex,* transl. and ed. G. V. Anrep.

Thompson, R. F. (2005). 'In search of memory traces'. *Annual Review of Psychology,* 56.

confabulation. 'Confabulation' as a technical term was first used by the German neurologists Bonhoeffer, Pick, and Wernicke in the early 1900s for false memory reports made by patients who suffered from a syndrome that later came to known as *Korsakoff's amnesia*. When asked what they did yesterday, these patients do not remember, but will report events that either did not happen, or happened long ago. During the remainder of the 20th century, the use of the term was gradually expanded to cover claims made by other types of patients, many of whom had no obvious memory problems, including patients who deny illness, *commissurotomy (split-brain) patients, patients with misidentification disorders (who make false claims about the identities of people), and patients with *schizophrenia, as well as children and normal adults in certain situations.

There are currently two schools of thought on the proper scope of the concept of confabulation, those who remain true to the original sense and so believe that the term should only be applied to false memory reports, and those who believe that the term can be usefully applied to a broader range of disorders. An examination of the etymology of the English word is not very helpful. When those German neurologists at the turn of the 20th century began using *'konfabulation',* they probably meant that their memory patients were creating fables when asked about their pasts. The patients were *fabulists*.

The technical definition of 'confabulation' the early neurologists coined has three components: (1) confabulations are false; (2) confabulations are reports; and (3) confabulations are about memories. There are significant problems with each of these three criteria, however. First, relying on falsity alone to characterize the problem in confabulation can produce arbitrary results. If a Korsakoff's syndrome patient, when asked what day of the week it is, happens to state correctly that it is Saturday, or an Anton's patient guesses correctly that the neurologist is holding up two fingers, we may still want to consider these to be confabulations. They are only true out of luck—simply made up, rather than the result of accurate tracking of the facts. Second, the idea that confabulations are reports, or stories, might be taken to imply that they are intrinsically linguistic in nature, in that they are always reports in the patient's natural language, such as German, and that hence confabulation is a strictly linguistic phenomenon. However, several researchers have categorized non-linguistic responses as confabulatory. One group had patients whose left hemispheres had been temporarily anaesthetized point to fabric samples with one hand to indicate which texture of fabric they had been stimulated with on the other hand. The patients also had the option of pointing to a question mark in trials in which they had not been stimulated, a non-linguistic version of answering 'I don't know'. Other researchers applied the term 'confabulation' to the behaviour of patients when they produced meaningless drawings as if they were familiar designs. Similarly, another group had patients reproduce from memory certain drawings they had seen, and referred to cases in which the patients added extra features to the drawings which were not actually present as confabulations. Finally, the problem with relating confabulations to memories is that, even in Korsakoff's syndrome, many confabulations are simply made up on the spot, and have little to do with any actual memories. That is, strictly speaking it is wrong to describe confabulations as memory reports. They are rather fictional fables, or at least false claims alleged to be memory reports.

Thus it seems that confabulations need not be false, may not be reports, and need not be about memories. If the original definition is problematic, that may be one reason why it was ignored by those who later described claims made by other, non-memory, patients. Patients who deny that they are paralysed have been claimed to confabulate when they provide reasons for why they cannot move ('My arthritis is bothering me', 'I'm tired of following your commands'). Another type of patient will deny blindness and attempt to answer questions about what he sees, producing what have been called confabulations ('It's too dark in here'). Misidentification patients have been said to confabulate when asked what the motives of the 'impostor' are, or why someone would go through all the trouble to impersonate someone else ('Perhaps my father paid him to take care of me'). Similarly, when the left hemispheres of split-brain patients attempt to answer questions without the necessary information (which is contained in their right hemispheres), this has also been called a confabulation.

This expansion forces several difficult questions about what had happened to the concept of confabulation. Has it expanded so much as to become meaningless? Do the new confabulation syndromes share anything significant with the classical memory cases? Some writers on confabulation have despaired of the fact that some of the confabulation syndromes involve memory (Korsakoff's, aneurysm of the anterior communicating artery), whereas others involve perception (denial of paralysis or blindness, split-brain syndrome, misidentification disorders). Since both memory and perception are knowledge domains, however, perhaps this indicates that the broader sense of 'confabulation' has to do with knowledge itself. According to this approach (Hirstein 2005), the brain's implementation of each knowledge domain—memory, perception, and introspection—is subject to characteristic confabulation syndromes.

1. Confabulations about memories

These are a defining characteristic of Korsakoff's syndrome and a similar syndrome caused by aneurysm of the anterior communicating artery (Kopelman 1987). Alzheimer's patients will often produce memory confabulations (see *dementia), and children up to a certain age are also prone to reporting false memories, apparently because their brain's prefrontal areas have not yet fully developed, while the Alzheimer's patients prefrontal lobes have been compromised by the amyloid plaque lesions. All of these confabulators have an initial memory retrieval problem, coupled with a failure to check and correct their false 'memories' (Johnson and Raye 1998). In contrast, there exist many memory patients with damage only to more posterior parts of the memory system (e.g. to the hippocampus or other parts of the temporal lobes) who freely admit that they cannot remember, and are not at all prone to producing confabulations (see *brain damage).

2. Confabulations about perceptions

Vision. Anton's syndrome patients are at least partially blind, but insist that they can see. Their posterior damage typically involves bilateral lesions to the occipital cortex, causing the blindness, coupled with prefrontal damage, causing the inability to become aware of the blindness. Split-brain patients will also confabulate when asked in certain situations about what they perceived.

Somatosensation. The patients who deny paralysis have a condition referred to as **anosognosia*, meaning unawareness of illness. They typically have a loss of one or more somatosensory systems for the affected limb. Apparently, certain types of damage (e.g. to the right inferior parietal lobe) can cause both the somatosensory problem, and at least temporarily affect prefrontal functioning enough to cause the confabulated denials of illness (Berti et al. 2005). The nature of the connections between frontal areas and the right inferior parietal lobe are less well understood. One possible connection is that the high level prefrontal executive processes based in the orbitomedial cortex are heavily dependent on the high level perceptual processing housed in the right inferior parietal lobe.

Person perception. Perceptual confabulations are also issued by patients suffering from the misidentification syndromes (especially Capgras syndrome). These syndromes may be caused by a deficit in representing the mind of the person who is misidentified (Hirstein 2008), coupled with an inability to realize the implausibility of the impostor claim.

3. Confabulations about introspection

Confabulations about intentions and actions. Patients who have undergone a *commissurotomy will tend to confabulate about actions performed by the right hemisphere. In a typical experiment, commands are sent to the right hemisphere only, but the left hemisphere, unaware of this, confabulates a reason for why the left hand obeyed the command. Similar sorts of confabulations can be elicited by brain stimulation. For example, the patient's cortex is stimulated, causing her arm to move. When asked why the arm moved, the patient claims she felt like stretching her arm. *Hypnotized people may also confabulate, e.g. the subject is given a hypnotic suggestion to perform a certain action, but then confabulates a different reason for it when asked.

There are many cases of confabulations about actions and intentions that do not involve the right hemisphere or any obvious lateral element (Wegner 2002). When Wilder Penfield electrically stimulated peoples' brains in the 1950s, he was able to cause them to make movements or emit sounds. Sometimes the patients would claim that Penfield was the cause of the movement. They responded with remarks such as, 'I didn't do that. You did' and, 'I didn't make that sound. You pulled it out of me' (Penfield 1975). In contrast, Hecaen et al. (1949) electrically stimulated a different area which caused the patients to perform'pill rolling' motions, or clench and unclench their fists. The patients claimed that they had done this intentionally, but were unable to offer a reason for the action. Delgado's brain stimulation patients also claimed they had performed the actions voluntarily, and confabulated a reason why. When Delgado (1969) stimulated yet another area, producing 'head turning and slow displacement of the body to either side with a well-oriented and apparently normal sequence, as if the patient were looking for something'. When the patients were asked why they engaged in those actions, genuine confabulations seemed to result:

> The interesting fact was that the patient considered the evoked activity spontaneous and always offered a reasonable explanation for it. When asked 'What are you doing?' the answers were, 'I am looking for my slippers,' 'I heard a noise,' 'I am restless,' and 'I was looking under the bed'. (Delgado 1969).

Confabulations about emotions. False attributions of emotions can count as confabulations. For example, in one experiment, people were given an injection of adrenaline (epinephrine) without their knowledge, but attributed their inability to sleep to, e.g., nervousness about what they had to do the next day. We may all be guilty of confabulating about our emotions on occasion, perhaps due to the combination of our feeling responsible for giving coherent accounts of our emotions and the opacity of our emotions to cognition.

Classifying confabulation syndromes as malfunctions in different knowledge domains eliminates the problem with the falsity criterion. The problem is not so much the falsity of their claims, it is rather their overall unreliability, at least in the affected domain. Confabulation seems to involve two phases of error. First, a flawed memory or response is created. Second, even with plenty of time to examine the situation and with urging from doctors and relatives, the patient fails to realize that the response is flawed. Our brains create flawed responses all the time. If I ask you if you have ever been inside the head of the Statue of Liberty, for instance, your brain is happy to provide an image of a view from inside, even if you've never been near the statue. But you are able to reject this as a real memory, so you catch the mistake at the second phase.

The brain processes capable of checking and correcting or rejecting flawed representations are called *executive processes*. Most executive processes reside in the prefrontal lobes, including the dorsolateral frontal lobes, on the side of the brain, the ventrolateral frontal lobes below them, and the orbitofrontal lobes, located just above the eye sockets (Rolls 1999, Fuster 2002). The following situations require the intervention of executive processes: planning or decision-making is required; there are no effective learned input–output links; a habitual response must be inhibited; an error must be corrected; the situation is dangerous; we need to switch between two or more tasks; or, we need to recall something. In theory, given the brain's large number of knowledge sources, there are many more confabulation syndromes than those listed here, but they should all follow the same pattern: damage to a knowledge system (either perceptual or mnemonic), typically located in the temporal or parietal lobes, coupled with damage to prefrontal executive processes responsible for monitoring and correcting the representations delivered by that epistemic system.

4. The locus of damage in confabulation

There are several clues as to the nature and location of the neurological damage in confabulation patients. (1) Confabulation about paralysis of the left arm can occur with stroke damage restricted to the right inferior parietal cortex. (2) The patients with aneurysms of the anterior communicating artery—a tiny artery near the anterior commissure that completes the anterior portion of the circle of Willis—provide our best clue about the locus of the frontal problems in memory confabulation (DeLuca and Diamond 1995). (3) Split-brain patients confabulate about information perceived by the right hemisphere. The right hemisphere, or lack of communication with the right hemisphere, shows up in all of the perceptual confabulations. Given the right hemisphere's greater role in producing and perceiving emotions, there may be a lateral element to the neural locus of confabulations about emotions. The cerebral commissures, the corpus callosum, and the anterior commissure are the three connecting fibre bundles between the two hemispheres. There are important functional links between the posterior orbitomedial cortex and the corpus callosum. Given the existence of dense interconnections between the left and right orbitomedial cortices, cutting their commissures may have the same effect of lesioning them directly.

5. The broader sense of 'confabulation'

The following definition is based on the idea that confabulation syndromes involve malfunctions in different knowledge domains, coupled with executive system damage (Hirstein 2005):

> Jan confabulates that *p* if and only if: (1) Jan claims that *p*. (2) Jan believes that *p*. (3) Jan's thought that *p* is ill-grounded. (4) Jan does not know that her thought is ill-grounded. (5) Jan should know that her thought is ill-grounded. (6) Jan is confident that *p*.

'Claiming' is broad enough to cover a wide variety of responses by subjects, including drawing and pointing. The second criterion captures the sincerity of confabulators. The third criterion refers to the problem that caused the flawed response to be generated. The fourth criterion refers to the failure of the second phase, the failure to reject the flawed response. The fifth criterion captures the normative element of our concept of confabulation. If the confabulator's brain was functioning properly, she would not make that claim. The last criterion refers to another important aspect of confabulators, the serene certainty they have in their communications, which may be connected to the frequent finding of low or abolished sympathetic autonomic activity in confabulating patients.

6. Confabulation and consciousness

Why does the anosognosic not notice what is missing? One message carried by the phenomena one encounters in a study of confabulation is that consciousness does not contain labels saying, 'an adequate representation of your left arm is missing' (denial); 'there is a gap in your memory here' (memory syndromes); 'you have no information about why your left arm just pointed at a picture of a cat' (split-brain syndrome); 'your representation of your father's mind is missing' (Capgras syndrome). The obvious hypothesis is that we confabulate because both the conscious data and the checker of that data are flawed.

Is confabulation then a type of *filling in, comparable to the way that the brain's visual system fills in the optic blind spot? Confabulation might be considered filling in at a higher, social level. It fills in social gaps in information: the doctor has asked for information, for example, so the patient supplies it. More sceptical writers seem to see consciousness itself as a massive confabulation, a *user illusion*. There may also be information here relevant to another question: What is the function of consciousness? The existence of confabulation supports the idea that consciousness functions as a testing ground, where thoughts and ideas can be checked, before they are allowed to become beliefs or participate in the causing of actions.

WILLIAM HIRSTEIN

Berti, A., Bottini, G., Gandola, M. et al. (2005). 'Shared cortical anatomy for motor awareness and motor control'. *Science,* 309.

Delgado, J. M. R. (1969). *Physical Control of the Mind: Toward a Psychocivilized Society*.

DeLuca, J. and Diamond, B. J. (1995). 'Aneurysm of the anterior communicating artery: a review of the neuroanatomical and neuropsychological sequelae'. *Journal of Clinical and Experimental Neuropsychology,* 17.

Fuster, J. (2002). *Cortex and Mind: Unifying Cognition*.

Hecaen, H., Talairach, J., David, M., and Dell, M. B. (1949). 'Coagulations limitées du thalamus dans les algies du syndrome thalamique: resultats thérapeutiques et physiologiques. *Revue Neurologique (Paris),* 81.

Hirstein, W. (2005). *Brain Fiction: Self-Deception and the Riddle of Confabulation*.

—— (2008). 'Confabulations about people and their limbs, present or absent'. In Bickle, J. (ed.) *Oxford Handbook of Philosophy and Neuroscience*.

Johnson, M. K., and Raye, C. L. (1998). 'False memories and confabulation'. *Trends in Cognitive Sciences,* 2.

Kopelman, M. D. (1987). 'Two types of confabulation'. *Journal of Neurology, Neurosurgery, and Psychiatry,* 50.

Penfield, W. (1975). *The Mystery of the Mind*.

Rolls, E. T. (1999). *The Brain and Emotion*.

Wegner, D. M. (2002). *The Illusion of Conscious Will*.

confidence judgement. See SIGNAL DETECTION THEORY

connectionist models. Connectionist models, also known as *parallel distributed processing* (PDP) models, are a class of computational models often used to model aspects of human perception, cognition, and behaviour, the learning processes underlying such behaviour, and the storage and retrieval of information from memory. The approach embodies a particular perspective in cognitive science, one that is based on the idea that our understanding of behaviour and of mental states should be informed and constrained by our knowledge of the neural processes that underpin cognition. While *neural network* modelling has a history dating back to the 1950s, it was only at the beginning of the 1980s that the approach gained widespread recognition, with the publication of two books (McClelland and Rumelhart 1986, Rumelhart and McClelland 1986), in which the basic principles of the approach were laid out, and its application to a number of psychological topics were developed. Connectionist models of cognitive processes have now been proposed in many different domains, ranging from different aspects of language processing to cognitive control, from perception to memory. The specific architecture of such models often differs substantially from one application to another, but all models share a number of central assumptions that collectively characterize the 'connectionist' approach in cognitive science. One of the central features of the approach is the emphasis it has placed on mechanisms of change. In contrast to traditional computational modelling methods in cognitive science, connectionism takes it that understanding the mechanisms involved in some cognitive process should be informed by the manner in which the system changed over time as it developed and learned. Understanding such mechanisms constitutes a significant part of current research in the domain (Elman et al. 1996; Mareschal et al. 2007a, 2007b).

Connectionist models take their inspiration from the manner in which information processing occurs in the brain. Processing involves the propagation of activation among simple units (artificial neurons) organized in networks, i.e. linked to each other through weighted connections representing synapses or groups thereof. Each unit then transmits its activation level to other units in the network by means of its connections to those units. The *activation function,* that is, the function that describes how each unit computes its activation based on its inputs, may be a simple linear function, but is more typically non-linear (e.g. a sigmoid function).

1. Representation, processing, and learning in connectionist networks
2. Connectionism and consciousness

1. Representation, processing, and learning in connectionist networks

Representation can take two very different forms in connectionist networks, neither of which corresponds to 'classical' propositional *representations. One form of representation is the pattern of activation over the units in the network. Units in some connectionist networks are specifically designated in advance by the modeller to represent specific items such as identifiable visual features, letters, words, objects, etc. Networks that employ such units for all cognizable entities of interest are called *localist* networks—the representation