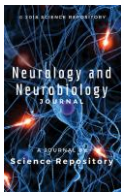

Available online at www.sciencerepository.org

Science Repository

Review Article

Towards a Neural Network Hypothesis for Functional (Dissociative) Amnesia: Catastrophic Forgetting

AJ Larner*

Honorary Senior Research Fellow, Department of Brain Repair & Rehabilitation, Institute of Neurology, University College London, UK

ARTICLE INFO

Article history:
Received: 15 July, 2022
Accepted: 2 August, 2022
Published: 19 August, 2022

Keywords:
Artificial neural networks
catastrophic forgetting
catastrophic interference
functional amnesia
functional cognitive disorders
neurology
psychiatry
sleep

ABSTRACT

Functional amnesia, also known as dissociative amnesia, psychogenic amnesia, or mnestic block syndrome, is a rare disorder which, although clinically heterogeneous, is most often characterised by dense retrograde amnesia mainly affecting the episodic-autobiographical domain but with relative preservation of anterograde memory function, a pattern dissimilar to that seen in other amnesic disorders. The pathogenesis of functional amnesia remains unknown. Here, appeal is made to the study of artificial neural networks in the hope that, as in other mnestic disorders, this might give insight into the mechanisms underpinning functional amnesia. Specifically, the observation of catastrophic forgetting or catastrophic interference occurring in artificial neural networks, that is the abrupt and complete loss of previously learned information when learning new information, is extended to the human nervous system to develop a novel hypothesis: the Catastrophic Forgetting Hypothesis of functional amnesia.

Introduction: Functional Amnesia

The disorder variously known as functional amnesia, dissociative amnesia, psychogenic amnesia, or mnestic block syndrome, has a heterogeneous phenotype but is typically characterised by dense retrograde amnesia mainly affecting the episodic-autobiographical domain of memory [1, 2]. The clinical retrograde amnesia may be of such density that even heavily over-learned material such as personal identity and personal semantic memory may be lost (this is the form of amnesia most often portrayed in popular films, no doubt because of its dramatic potential, however clinically implausible the scenario may be [3, 4]). One suggested definition of functional amnesia is the inability to recall autobiographical information consciously in the absence of structural neuroimaging evidence of brain damage [1].

Patients affected with this disorder tend to be young, usually in their 20s-40s, often with a past history of psychiatric disease (depression), and with evidence for a triggering event such as psychological stress, or physical trauma, the latter often mild. The retrograde amnesia typically shows a temporal gradient in which more recent events are better recalled than distant ones. The ability to learn new information is usually preserved, hence there is generally no anterograde amnesia (exceptions may occur). Functional neuroimaging may show changes in networks subserving autobiographical memory in right frontotemporal regions. Recovery of "lost" memories may sometimes occur, partially or totally, after variable time periods, but disability may be long-lasting [1, 2].

This disorder is included in the most recent (2013) edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-5) of the American Psychiatric Association as "dissociative amnesia" and is listed amongst the dissociative disorders, as 300.12 (F44.0). However, it is also included in the heterogeneous category of "functional cognitive disorders" (FCD) proposed by Stone *et al*., along with various other possible typologies (mood disorder, other functional disorders, medication effects, dementia health anxiety, normal cognitive experience, malingering) [5]. This categorisation has gained widespread acceptance in recent years and has been instrumental in promoting the

**Correspondence to: AJ Larner, MD, DHMSA, PhD, FRCP(UK), Honorary Senior Research Fellow, Department of Brain Repair & Rehabilitation, Institute of Neurology, University College London, UK; E-mail: andrew.larner2@nhs.net*

http://dx.doi.org/10.31487/j.NNB.2022.03.02

development of operational diagnostic criteria for FCD [6, 7]. Hence the term "functional amnesia" is used here.

The pattern of memory dysfunction in functional amnesia differs from that seen in other mnestic disorders. Specifically, the temporal gradient of the retrograde amnesia is a reversal of that seen in Alzheimer's disease (AD) and transient global amnesia (TGA) wherein more recent memories are typically impaired (Ribot's law). This Ribot temporal gradient is thought to relate to effective disconnection (AD) or temporary functional ablation (TGA) of the hippocampus, which is also critical for the ability to learn new information which is progressively (AD) or transiently (TGA) lost.

The different terminologies which have been used to describe this condition reflect uncertainty about its pathogenesis. Existing hypotheses implicate an "inability of access" to autobiographical memories, possibly as a consequence of a stressful (psychological or physical) event which triggers neuroendocrine changes (hypercortisolaemia, changes in the hypothalamo-pituitary axis) which block retrieval of stress-related memories with generalisation to other memories [1]. This may be interpreted as a mechanism of protection against an adverse environment, occurring in individuals predisposed by prior stress or trauma ("two-hit hypothesis") [1]. Recovery occurs if or when a pathway to autobiographical memories is "unblocked" and access is restored.

Appeal to the study of artificial neural networks may give insight into the mechanisms of brain function and dysfunction, including psychiatric disorders, and disorders of memory such as TGA and FCD [8-10]. Here, a neural network hypothesis for functional amnesia is proposed based on a phenomenon encountered during sequential learning in artificial neural networks: catastrophic forgetting or catastrophic interference.

Catastrophic Forgetting or Catastrophic Interference

Human learning is characterised by the continuous, sequential assimilation of new information incrementally. In contrast, artificial neural networks show a tendency to forget previously learned information completely and abruptly when learning new information, a phenomenon known as catastrophic forgetting or catastrophic interference [11, 12]. In connectionist models, this behaviour is often framed mechanistically as a trade-off between the competing requirements of the stability of connections for memory retention and the plasticity of connections for new learning. Methods to overcome catastrophic forgetting have been suggested, essentially changing the internal structure of the neural network, such as "freezing" subsets of connections ("synapses") which mediate previous learning ("old memories") [13].

Catastrophic forgetting or interference behaviour has been considered not to occur in biological, as opposed to artificial, nervous systems. It may be inferred that there would be strong evolutionary pressure for learning capacity of nervous systems to be characterised by continuous assimilation of new information, rather than by vulnerability to abrupt and complete forgetting. Hence various biological mechanisms have been considered to prevent catastrophic forgetting. For example, the observation of neurogenesis in the adult hippocampus, a phenomenon not encountered elsewhere in the mammalian central nervous system, has been suggested to be a mechanism to avoid catastrophic interference [14]. Likewise, sleep may be a mechanism which, at least in part, serves to protect old memories from being forgotten after new learning [15]. The process of pattern separation, whereby overlapping or very similar inputs to the hippocampus are separated from one another by enacting a sparse coding scheme, might also reduce the possibility of catastrophic interference [16].

Failure or impairment of these mechanisms might render the brain vulnerable to catastrophic forgetting with predictable clinical consequences: functional amnesia.

Hypothesis: The Catastrophic Forgetting Hypothesis of Functional Amnesia

The hypothesis proposed is that the process of catastrophic forgetting or interference, as encountered in artificial neural networks when attempting sequential learning tasks, may occur in the human nervous system in particular circumstances, resulting in the specific condition of functional amnesia. The proposed chain of causation is as follows. An event, be it psychological or physical trauma, occurring in an individual with predisposing psycho-socio-biological factors, triggers an episode of catastrophic forgetting which manifests as the syndrome of functional amnesia.

Specifically, as a consequence of catastrophic forgetting, there is apparent loss (or loss of access to) previously learned information, manifest clinically as retrograde amnesia, just as in artificial neural networks undergoing catastrophic forgetting or interference there is abrupt and complete loss of previously learned information. This may include loss of personal identity and other autobiographical information.

Despite this profound retrograde amnesia, patients with functional amnesia retain the ability to learn new information, that is they do not have anterograde amnesia. This finding would be predicted as a further consequence of catastrophic forgetting, since in artificial neural networks the learning of new information occurs despite the loss of previously learned information. Patients with functional amnesia may be able to "relearn" forgotten autobiographical information following the onset of their disorder. Again, this clinical observation might be predicted as a consequence of catastrophic forgetting, since artificial neural networks may require retraining on previously learned inputs, or reactivation of previous inputs, to prevent new learning from interfering with previously learned information.

Hence a process of catastrophic forgetting or interference might explain many of the clinical observations made in patients with functional amnesia, by analogy with the consequences of catastrophic forgetting in artificial neural networks.

Testable Predictions of the Catastrophic Forgetting Hypothesis of Functional Amnesia

Unlike other forms of FCD, which are commonly observed in day-to-day clinical practice, cases of functional amnesia are relatively rare [17]. For example, one tertiary center with a specialist interest in functional amnesia reported 53 cases seen over a period of nearly 20 years [2]. The

paucity of case material is one factor limiting potential testing of this hypothesis. Moreover, as in other amnesic syndromes, patients with functional amnesia may reach clinical attention only after the most significant pathophysiological neural events have occurred, all investigational findings then reflecting changes downstream from the initiating events [9]. Nevertheless, some possible tests of the hypothesis may be suggested.

Sleep is recognised to have a critical role in memory consolidation. The "Overfitted Brain Hypothesis" proposed that the evolved function of dreaming is as a protective mechanism against overfitting, and on the basis of this hypothesis it has been postulated that sleep disturbance, specifically of dreaming, might have a role in FCD [8, 18]. Implementing sleep-like phases in artificial neural networks has been shown to reduce catastrophic forgetting [15]. If sleep is a protective factor against catastrophic forgetting in biological systems, then examination of sleep architecture in functional amnesia patients might be predicted to show changes predisposing to catastrophic forgetting. Sleep disturbance in functional amnesia would not be a surprising finding in the context of any underlying depression, wherein sleep is typically disturbed. Moreover, recognised precipitating events for functional amnesia (psychological or physical trauma) might also impact on sleep dynamics. Remediation of any sleep deficit has been suggested as an adjunctive approach in the treatment of functional amnesia [1].

Biological systems have been thought to avoid catastrophic forgetting by various mechanisms. If this general principle were contradicted, by showing that catastrophic forgetting could occur in the human brain, this would provide some indirect support for the Catastrophic Forgetting Hypothesis of functional amnesia. Experimental studies have been reported which suggest that susceptibility to catastrophic interference can be demonstrated in the human brain using "Fast Mapping", an incidental exclusion-based learning mechanism that supports hippocampal-independent learning [19]. This methodology might also be applied to recovered or recovering functional amnesia patients to see if they are also susceptible to catastrophic forgetting. Although there are obviously some delicate ethical issues to be navigated in any such investigation (although the methodology has been applied to amnesic patients), this might provide the most direct evidence to support the hypothesis of a role for catastrophic interference in functional amnesia [19].

Limitations of the Catastrophic Forgetting Hypothesis of Functional Amnesia

Aside from the aforementioned difficulties in testing the Catastrophic Forgetting Hypothesis of functional amnesia, there are additional limitations to consider. Functional amnesia appears to be heterogeneous at the clinical level so a single, over-arching hypothesis which proposes to explain all instances of this entity is unwarranted [1, 2].

Functional amnesia patients sometimes regain "lost" memories, partially or completely, sometimes years after the onset of their disorder [1, 2]. As implied by the term "mnestic block syndrome", the problem may therefore be one of access to, rather than complete loss of, memories, at least in some cases. Such an outcome would not necessarily be predicted by simple extrapolation from the catastrophic forgetting behaviour seen in artificial neural networks. This divergence may simply be a reflection of the greater complexity of the human brain in comparison to artificial neural networks, with mechanisms instantiated to prevent catastrophic forgetting. To use Hughlings Jackson's pathophysiological framework, the negative features of functional amnesia may reflect non-functioning rather than destroyed tissue. Apparent recovery of "lost" memories might also suggest lack of metacognition plays a role in at least some instances of functional amnesia, as has previously been suggested in other forms of FCD [20-24].

The hypothesis as proposed obviously lacks neuroanatomical and mechanistic precision, a complaint frequently levelled against hypothetical models of functional disorders [22, 25]. Ideally studies of hippocampal and neocortical networks pertinent to memory are required in functional amnesia patients. The reversed temporal gradient of the retrograde amnesia may implicate primarily neocortical rather than hippocampal mechanisms. Although memory acquisition is dependent on the hippocampus, the limited capacity of this system may be one factor requiring transfer of information to the neocortex for the purposes of consolidation, information for older memories becoming hippocampus-independent (and more semantic) over time.

Mechanistically, memory retrieval is thought to involve reinstatement in different cortical areas of activity present during the learning of an episode. This heteroassociation contrasts with the autoassociation in the hippocampal CA3 region recurrent collateral networks which is thought to be necessary for new learning. Attractor networks, based on the cortical anatomy of recurrent collateral excitatory synaptic connections between pyramidal neurons, may constitute a fundamental principle of cerebral cortical function [26]. Hippocampus CA3 may be characterised as a single global autoassociative attractor network, catastrophic collapse in the function of which has been suggested to result in TGA [9]. Neocortex has been modelled as multiple local discrete and continuous attractors, based on the recurrent collateral connections of neocortical pyramidal neurons in both superficial and deep cortical layers (2/3 and 5 respectively). Catastrophic degradation in the function of these attractors might be the mechanism of catastrophic forgetting, differential involvement accounting for the heterogeneity of the clinical features of functional amnesia.

Discussion

A reciprocal relationship between studies of human neuroscience and of artificial neural networks is acknowledged, each discipline potentially capable of informing the other. For example, sleep as a protection against catastrophic forgetting was explored in artificial neural networks based on the known role of sleep in the consolidation of human memory [15].

Deep learning by artificial neural networks is subject to (at least) two serious issues that have been considered not to occur in natural nervous systems: overfitting and catastrophic forgetting [27]. A previous hypothesis proposed a role for overfitting in the pathogenesis of FCD [10]. The current hypothesis proposes a role for catastrophic forgetting in functional amnesia. This is based on the correspondences between the phenomena of catastrophic forgetting (abrupt loss of previously learned information with preserved new learning ability) and the clinical observations in functional amnesia (dense retrograde amnesia, preserved

anterograde memory function). The evolved functions of sleep and dreaming may thus be pertinent to the prevention of both overfitting and catastrophic forgetting in the human brain [10, 18].

REFERENCES

1. Markowitsch HJ, Staniloiu A (2016) Functional (dissociative) retrograde amnesia. *Handb Clin Neurol*. 139: 419-445. [Crossref]
2. Harrison NA, Johnston K, Corno F, Casey SJ, Friedner K et al. (2017) Psychogenic amnesia: syndromes, outcome, and patterns of retrograde amnesia. *Brain* 140: 2498-2510. [Crossref]
3. Baxendale S (2004) Memories aren't made of these: amnesia at the movies. *BMJ* 329: 1480-1483. [Crossref]
4. Larner AJ (2008) Neurological literature: cognitive disorders. *Adv Clin Neurosci Rehab* 8: 20.
5. Stone J, Pal S, Blackburn D, Reuber M, Thekkumpurath P et al. (2015) Functional (psychogenic) cognitive disorders: a perspective from the neurology clinic. *J Alzheimers Dis* 48: S5-S17. [Crossref]
6. Ball HA, McWhirter L, Ballard C, Bhome R, Blackburn DJ et al. (2020) Functional cognitive disorder: dementia's blind spot. *Brain* 143: 2895-2903. [Crossref]
7. Ball HA, McWhirter L, Ballard C, Bhome R, Blackburn DJ et al. (2021) Reply: Functional cognitive disorder: dementia's blind spot. *Brain* 144: e73. [Crossref]
8. Durstewitz D, Koppe G, Meyer Lindenberg A (2019) Deep neural networks in psychiatry. *Mol Psychiatry* 24: 1583-1598. [Crossref]
9. Larner AJ (2022) Transient global amnesia: model, mechanism, hypothesis. *Cortex* 149: 137-147. [Crossref]
10. Larner AJ (2022) Towards a neural network hypothesis for Functional Cognitive Disorders: an extension of the Overfitted Brain Hypothesis. *Cogn Neuropsychiatry* 27: 314-321. [Crossref]
11. McCloskey M, Cohen NJ (1989) Catastrophic interference in connectionist networks: the sequential learning problem. In: Bower, G.H. (ed.). Psychology of Learning and Motivation. Volume 24. New York: Academic Press, 109-165.
12. Ratcliff R (1990) Connectionist models of recognition memory: constraints imposed by learning and forgetting functions. *Psychol Rev* 97: 285-308. [Crossref]
13. Kirkpatrick J, Pascanu R, Rabinowitz N, Veness J, Desjardins G et al. (2017) Overcoming catastrophic forgetting in neural networks. *Proc Natl Acad Sci USA* 114: 3521-3526. [Crossref]
14. Wiskott L, Rasch MJ, Kempermann G (2006) A functional hypothesis for adult hippocampal neurogenesis: avoidance of catastrophic interference in the dentate gyrus. *Hippcampus* 16: 329-343. [Crossref]
15. González OC, Sokolov Y, Krishnan GP, Delanois JE, Bazhenov M (2020) Can sleep protect memories from catastrophic forgetting? *Elife* 9: e51005. [Crossref]
16. Knierim JJ, Neunuebel JP (2016) Tracking the flow of hippocampal computation: pattern separation, pattern completion, and attractor dynamics. *Neurobiol Learn Mem* 129: 38-49. [Crossref]
17. Bharambe V, Larner AJ (2018) Functional cognitive disorders: demographic and clinical features contribute to a positive diagnosis. *Neurodegener Dis Manag* 8: 377-383. [Crossref]
18. Hoel E (2021) The overfitted brain: dreams evolved to assist generalization. *Patterns (N Y)* 2: 100244. [Crossref]
19. Merhav M, Karni A, Gilboa A (2014) Neocortical catastrophic interference in healthy and amnesic adults: a paradoxical matter of time. *Hippocampus* 24: 1653-1662. [Crossref]
20. Larner AJ (2018) Dementia screening: a different proposal. *Future Neurol* 13: 177-179.
21. Bhome R, McWilliams A, Huntley JD, Fleming SM, Howard RJ (2019) Metacognition in functional cognitive disorder – a potential mechanism and treatment target. *Cogn Neuropsychiatry* 24: 311-321. [Crossref]
22. Larner AJ (2021) Functional cognitive disorders (FCD): how is metacognition involved? *Brain Sci* 11: 1082. [Crossref]
23. Bhome R, McWilliams A, Price G, Poole NA, Howard RJ et al. (2022) Metacognition in functional cognitive disorder. *Brain Commun* 4: fcac041. [Crossref]
24. Larner AJ (2022) Metacognition in functional cognitive disorder: contradictory or convergent experimental results? *Brain Commun* 4: fcac138. [Crossref]
25. Brown RJ, Reuber M (2016) Towards an integrative theory of psychogenic non-epileptic seizures (PNES). *Clin Psychol Rev* 47: 55-70. [Crossref]
26. Rolls ET (2016) Cerebral cortex. Principles of operation. Oxford: Oxford University Press.
27. Xie Z, He F, Fu S, Sato I, Tao D et al. (2021) Artificial neural variability for deep learning: on overfitting, noise memorization, and catastrophic forgetting. *Neural Comput* 33: 2163-2192. [Crossref]