Armin Schnider
The Confabulating Mind: How The Brain Creates Reality.
New York: Oxford University Press 2008.
341 pages
US$59.95 (cloth ISBN 978-0-19-920675-9)

Written by Armin Schnider, a professor of Neurorehabilitation at the University Hospital of Geneva, this book contains a comprehensive examination of a fascinating phenomenon of confabulation, an unintentional false production from memory. Starting with a historical review of early clinical studies and observations, the author proceeds to examine current studies of the phenomenon, its various aetiologies and proposed mechanisms, and presents a neurophysiological explanatory hypothesis, drawing on specialized controlled group studies and the results of PET and fMRI imaging. As the title suggests, the ultimate goal of the research presented in the book is to gain insight into the correct production from memory in a healthy brain and its role in shaping the mind's picture of reality.

The descriptions of confabulation, which emerged in the late eighteen hundreds in connection with the Korsakoff Syndrome, varied not only in classifications but also in aetiologies and proposed mechanisms. The classifications and explanations arising from these early studies, grounded mostly in clinical observation, were often limited by the aetiologies of the disorders in the observed patients. While some authors proposed a mechanism based on memory gap-filling, others emphasized the temporal order deficit; still others insisted that the type of personality (introversion, inferiority) was also an essential factor. The association with Korsakoff Syndrome influenced the general acceptance of the toxic origin of the phenomenon, until the advent of imaging technologies shed the light on the role of focal damage in the area of ventromedial frontal lobe.

The early classifications distinguished momentary confabulations—provoked, plausible accounts allegedly produced out of embarrassment to fill a memory gap—and fantastic confabulations—fabrications related to unreal, imaginary events. Citing their own clinical observations and controlled studies using California Verbal Learning Test, Schnider revises the classification, distinguishing between four forms of confabulation: intrusions in memory tests (or simple 'provoked confabulations'); 'momentary confabulations' (false statements in response to incitement); 'fantastic confabulations' (often illogical, with no basis in reality); and 'behaviorally spontaneous confabulations' (confusion of reality underlying inappropriate behaviour and statements).

Provoked confabulations—intrusions in memory tests similar to intrusions with healthy subjects over longer delay—appear to be non-specific signs of brain dysfunction with no anatomical specificity. Fantastic confabulations, on the other hand, had been described in patients with advanced paralytic dementia, schizophrenia, psychosis, and in the acute stages of traumatic brain injury or ACoA aneurism rapture. Importantly,

elaborate momentary confabulations often involve damage to ventromedial prefrontal lobes and the diencephalon, while the behaviorally spontaneous confabulations involve damage to the posterior orbitofrontal cortex or related anterior limbic structures.

In the second half of the twentieth century, the term 'confabulation' had been extended to other disorders characterized by false statements, such as anosognosia, false recognition, or paramnesic misidentification (déjà-vu, Capgras, Fregoli, intermetamorphosis, reduplicative paramnesia). According to Schnider, this semantic congruence has led some researchers to postulate common mechanisms for both mnestic and non-mnestic confabulations. Although behaviorally spontaneous confabulations might be accompanied by false recognition and reduplicative paramnesia (for place and person), which suggests a common neurological basis, there is no consistent association between mnestic and non-mnestic confabulations such as anosognosia. Furthermore, Schnider dismisses the existence of a common mechanism between severe pathological confabulation and the production of false memories in healthy subjects through suggestive questioning or manipulation of post-event processing. Normal false memories require manipulation and appear to be generated during the encoding and re-encoding rather than the retrieval of memories.

In discussing proposed mechanisms, Schnider expresses doubt that a common mechanism underlying such diverse (anatomically, clinically, and pathologically) forms of confabulations can be found. Controlled data from groups of patients show that confabulators neither require a gap in memory nor exhibit an increased tendency to fill such gaps. Schnider is equally doubtful about the role of motivation and personality as a mechanism of confabulation: rather than being the cause of confabulations, positive emotional bias, if it exists, might be the consequence of the anterior limbic lesions affecting the brain's reward system, which are common in spontaneous behavioral confabulators. The executive failure theory is also unconvincing as a sufficiently specific causal account rather than simply a concurrent effect of general cognitive impairment. Finally, although hypotheses based on defective memory retrieval and monitoring process appear to fit case observations, no physiological mechanism or specific monitoring deficit had been experimentally demonstrated.

Focusing exclusively on spontaneous behavioral confabulation and drawing on a variety of experimental results, Schnider presents an account of such physiological mechanism and monitoring deficit. The confusion of reality typical of behavioral confabulation is due to the failure to distinguish between memories that pertain to on-going reality and those that do not—failure of the memory filtering mechanism to suppress currently irrelevant memories. Lesion and functional imaging data suggest that damage to the anterior limbic structure or the posterior medial orbitofrontal cortex, specifically area 13 and the ventromedial cortex, is critical to the failure of filtering. In healthy brains, the filtration, inducing bilateral activity in area 13 of the orbitofrontal cortex, occurs 200-300 ms after stimulus presentation, preceding the stage of recognition

and re-encoding (400-600 ms). This means that the brain sorts out the currently irrelevant memories before they are consciously recognized and leaves them out of the processing stage characterized by widespread neocortical synchronization. The orbitofrontal memory filtering mechanism, responsible for the adaptation of thought and behaviour to reality, is pre-conscious.

Yet, the failure of pre-conscious memory filtering does not explain why, at a conscious level, patients with spontaneous behavioral confabulation fail to recognize blatant inconsistencies between their current situation and recalled memories. Schnider speculates that this failure is due to a breakdown in reality checking and extinction, which he connects to the posterior orbitofrontal area and the brain's reward system. The posterior orbitofrontal cortex is characterized by two networks, medial and orbital, with strong, distinct connections to the amygdala and hypothalamus. These networks have separate projections upon the ventral striatum and pallidum, which converge on the dopaminergic part of the substantia nigra. Thus, the ventromedial prefrontal networks of the posterior orbitofrontal area have connections with the brain's reward system.

In learning tasks, cues invoke memories and induce mental associations, leading to anticipation of future outcomes. The reality check involves comparing the anticipated outcome to the actual outcome, and adapting behavior and thinking to the results. The reward system plays an important role in reality checking and learning not only by arousing expectations and reinforcing associations but also in the extinction of such associations when expectations are no longer satisfied. Schnider suggests that the failure of behavioral confabulators to integrate the absence of anticipated outcomes might be viewed as corresponding to the failure of extinction, which is well known from animal experiments. He reports the results of PET and electrophysiological studies in support of the view that the orbitofrontal cortex may function as a generic outcome monitoring system.

Citing independent animal experiments, Schnider suggests that damage to the posterior medial orbitofrontal area, specifically area 13, impairs extinction and creates deficit in learning new stimulus-reward associations contradicting previously established ones. Furthermore, single cell response studies in animals identified a type of cell, located predominantly in area 13, that fires to signal deviation from reward expectancy. This led Schnider to speculate that the firing of these neurons in the orbitofrontal cortex transiently inhibits the activity of dopaminergic neurons in the brain's reward system, known to signal non-deliverance of reward.

Schnider's book is impressive in its breadth, focus, and collected evidence. But does it really live up to its name—*The Confabulating Mind; How the Brain Creates Reality*? Although a large part of the book deals with the historical and current overview of literature on confabulation, it is perhaps too broad and, as a result, too cursory. The treatment of such curious phenomena as paramnesic misidentification, non-mnestic

confabulations, and provoked confabulations in healthy subjects is brief and leaves the impression that the author, dismissing the phenomena as confabulation only in virtue of 'semantic congruence', is eager to move on to the discussion of what he finds really interesting—the spontaneous behavioral confabulation. Here, the author has much to say, but this discussion, leaving out many curious aspects of confabulation, can claim to explain only a part of the confabulating mind that we find so fascinating.

Still, the small part that can be explained yields amazing results: not only does our picture of reality depend on filtering out irrelevant memories, but this online filtering is pre-conscious! This raises a number of interesting questions: What is it that allows the orbitofrontal cortex to recognize a relevant memory in normal processing? What does the fact that the irrelevant memories, which seep through malfunctioning filtering, often reflect old habitual routines, tell us about cognitive architecture and recall? What is the relationship between the strength of association among memories, the order of recall, and the recognition of mnestic relevance? What role, if any, does memory encoding process play in how the orbitofrontal filtering mechanism determines relevance upon recall? Schnider's work might not explain how the brain creates reality until a fuller causal account of the filtering mechanism is provided, but it is certainly an excellent start in this direction.

Anton Petrenko
York University