CORTEX 87 (2017) 5–15

Available online at www.sciencedirect.com

ScienceDirect

Journal homepage: www.elsevier.com/locate/cortex

Special issue: Review

What does extinction have to do with confabulation?

Armin Schnider [a,b,*], *Louis Nahum* [a,b] *and Radek Ptak* [a,b]

[a] *Laboratory of Cognitive Neurorehabilitation, Division of Neurorehabilitation, University of Geneva, Geneva, Switzerland*

[b] *Division of Neurorehabilitation, Department of Clinical Neurosciences, University Hospital of Geneva, Geneva, Switzerland*

ARTICLE INFO

Article history:
Received 5 April 2016
Reviewed 30 May 2016
Revised 31 July 2016
Accepted 21 October 2016
Published online 1 November 2016

Keywords:
Korsakoff syndrome
Confabulation
Schizophrenia
Orbitofrontal cortex
Behavioural extinction
Decision making

ABSTRACT

Behaviourally spontaneous confabulation denotes a distinct syndrome consisting of confabulations that patients act upon, disorientation, and amnesia. It corresponds to the stable form of the original Korsakoff syndrome. While the syndrome may also occur in confusional states and degenerative dementia, this article is about the syndrome as it occurs after acute and focal brain damage. The patients act according to ideas and obligations that can mostly be traced back to real experiences in their past, but which are not currently valid guides of thinking and behaviour. This inability to abandon behavioural guides (anticipations) that are currently not valid corresponds to a failure of behavioural extinction and to the inability to abandon a previously rewarded choice in reversal learning when the expected reward (outcome) fails to occur, that is, following extinction trials. This article describes evidence from human and animal experiments showing that the posterior medial orbitofrontal cortex (OFC), which is typically damaged in these patients, and connected structures of the reward system contain the neural apparatus to signal the non-occurrence of anticipated outcomes, thereby presumably synchronizing thought and behaviour with current reality. Failure of this function, which we call orbitofrontal reality filtering, is associated with behaviourally spontaneous confabulation and disorientation after acute and focal brain damage, but not with other forms of confabulation, and not with reality confusion in degenerative dementia. Potential links with psychosis and decision making will be discussed.

1. Introduction

The term confabulation was proposed by Wernicke (1900) to denote "the emergence of memories of events and experiences which never took place". It replaced Korsakoff's earlier term "pseudo-reminiscences", defined as the "situation in which a patient conceives of an event that he has not really experienced, but that has only come to his mind, as if it had really happened to him" (Korsakoff, 1891). Confabulations

* *Corresponding author*. Service de Neurorééducation, Hôpitaux Universitaires de Genève, Av. de Beau-Séjour 26, CH-1206 Geneva, Switzerland.
E-mail address: armin.schnider@hcuge.ch (A. Schnider).

http://dx.doi.org/10.1016/j.cortex.2016.10.015
0010-9452/

were an element of the original Korsakoff syndrome, as defined by Bonhoeffer (1901), which encompassed severe anterograde amnesia, temporally limited retrograde amnesia, disorientation and "a strong tendency to confabulation".

Bonhoeffer (1901) already distinguished two forms of confabulation: (1) Momentary or "out-of-embarrassment" confabulations produced to hide a gap in memory, and (2) fantastic confabulations that go beyond the need to fill a gap in memory. More detailed classifications have been proposed (Moll, 1915). Based on our own observations, Schnider (2008) has proposed to distinguish four forms of memory-related confabulations, which partially or fully dissociate from each other in terms of mechanism and anatomy (Schnider, 2008; Nahum, Bouzerda-Wahlen, Guggisberg, Ptak, & Schnider, 2012):

(1) Intrusions in memory tests (simple provoked confabulations). These are independent from other confabulations and disorientation and tend to correlate with correctly retrieved items. They seem to be the price for retrieval of more information than is actually available in memory.
(2) Momentary confabulations, that is, false verbal statements that patients make in response to questions or in a situation provoking a comment. These confabulations can be explored with questionnaires. There is no definite mechanism or anatomical basis. They may be associated with impaired mental flexibility, a tendency to fill gaps in memory, or they may be part of the following form of confabulation (Nahum et al., 2012).
(3) Behaviourally spontaneous confabulations, which occur in the context of amnesia and reflect a confusion of reality: patients are disoriented when asked about present circumstances (time, location, current duties etc.) and at least intermittently act according to their confabulations. This syndrome corresponds to the stable form of the originally described Korsakoff syndrome (Bonhoeffer, 1901), that is, when it occurs in patients who are no longer in a confusional state. This article is about this form of confabulation as it occurs after acute and focal brain damage.
(4) Fantastic confabulations, which are illogical, often nonsensical and incoherent ideas as they can be observed in confusional states, decompensated psychosis and advanced dementia.

What would reality confusion, as evident from confabulations that patients act upon and disorientation, have to do with behavioural extinction (called extinction in the following)? Extinction denotes a learning process in which the repeated non-occurrence of an anticipated outcome (e.g., a reward) leads to the gradual abandonment of the anticipation and the associated behaviour (Pavlov, 1927). Reward denotes a situation or a stimulus, which reinforces behaviour (Rolls, 1999). The reward system is a loosely defined brain system that processes the type and the occurrence (or non-occurrence) of reinforcing outcomes. While reward processing is often portrayed as quasi-equivalent with pleasure and emotion (Grabenhorst & Rolls, 2011), we will present evidence that the "reward system" also processes outcomes that have no tangible pleasure (hedonic) value. Core anatomical structures of the system are the orbitofrontal cortex (OFC), subcortical dopaminergic structures (substantia nigra, striatum), and the amygdala (Grabenhorst & Rolls, 2011). Of particular interest here will be the OFC, the dopaminergic structures, and their connections. This system not only processes the occurrence of rewards (outcomes), but also their non-occurrence (Schultz, Dayan, & Montague, 1997). Of particular interest for this paper will be the neural activities, which signal the non-occurrence of expected outcomes. We will use the term extinction capacity for the ability to integrate the non-occurrence of anticipated outcomes at specific points in time into the concept of current reality and to adapt behaviour accordingly. Thus, the term is used here to describe the signalling of singular events as it is necessary for rapid behavioural adaptation in reversal learning, rather than the learning process by repeated exposure which characterizes full behavioural extinction. We will provide evidence that the capacity to use this signal is necessary for humans to synchronize their thinking and behaviour with current reality ("the now") and that its failure induces reality confusion with confabulation and disorientation.

2. The clinical syndrome of behaviourally spontaneous confabulation

In most cases, behaviourally spontaneous confabulation firstly manifests as confabulations in discussions. Patients recount recent events and experiences as well as plans for the day, which do not correspond to reality and which may appear completely invented. Very often, however, these ideas can be traced back to real events and activities in the patient's past (Korsakoff, 1891; Schnider, von Däniken, & Gutbrod, 1996a). Most notably, and in distinction from "simple" momentary confabulations (form 2, above), the patients also intermittently act according to these confabulations, as if their ideas related to the current situation. Some examples may clarify the picture:

A 58-year old woman had suffered damage to the right posterior medial OFC following rupture of an aneurysm of the anterior communicating artery. While she did not confabulate at all in memory test (no intrusions) she profoundly misinterpreted her current situation. She stood up during a therapy session indicating that she had to feed her baby, who was over 30 years old at the time, and then desperately searched for him (Schnider et al., 1996a).

False ideas may be stable for a long period. A 63-year old psychiatrist, who had suffered primarily right posterior medial orbitofrontal damage after rupture of an aneurysm of the anterior communicating artery, was convinced for months that she was actually a staff psychiatrist in our unit. She would leave therapy sessions in the conviction that she had to see patients. She maintained this idea for several months, in addition to a second idea, namely, that she had to organize a reception in the evening. She slapped her husband when she found the fridge empty, convinced that he had hidden the food she had already bought for tonight's reception. She had indeed worked as a psychiatrist until an early retirement 15 years ago and had often arranged big receptions