Progress in Neurobiology 92 (2010) 345–369

Contents lists available at ScienceDirect

Progress in Neurobiology

journal homepage: www.elsevier.com/locate/pneurobio

Toward a neurobiology of delusions

P.R. Corlett [a,*], J.R. Taylor [a], X.-J. Wang [b], P.C. Fletcher [c], J.H. Krystal [a]

[a] Department of Psychiatry, Yale University School of Medicine, Connecticut Mental Health Centre, Abraham Ribicoff Research Facility, 34 Park Street, New Haven, CT 06519, USA
[b] Department of Neurobiology and Kavli Institute of Neuroscience, Yale University School of Medicine, New Haven, CT, USA
[c] Department of Psychiatry, Brain Mapping Unit and Behavioural and Clinical Neurosciences Institute, University of Cambridge, School of Clinical Medicine, Cambridge,

ARTICLE INFO

Article history:
Received 11 February 2010
Received in revised form 6 May 2010
Accepted 8 June 2010

Keywords:
Delusions
Prediction
Error
Learning
Memory
Reconsolidation
Habit

United Kingdom

ABSTRACT

Delusions are the false and often incorrigible beliefs that can cause severe suffering in mental illness. We cannot yet explain them in terms of underlying neurobiological abnormalities. However, by drawing on recent advances in the biological, computational and psychological processes of reinforcement learning, memory, and perception it may be feasible to account for delusions in terms of cognition and brain function. The account focuses on a particular parameter, prediction error – the mismatch between expectation and experience – that provides a computational mechanism common to cortical hierarchies, fronto-striatal circuits and the amygdala as well as parietal cortices. We suggest that delusions result from aberrations in how brain circuits specify hierarchical predictions, and how they compute and respond to prediction errors. Defects in these fundamental brain mechanisms can vitiate perception, memory, bodily agency and social learning such that individuals with delusions experience an internal and external world that healthy individuals would find difficult to comprehend. The present model attempts to provide a framework through which we can build a mechanistic and translational understanding of these puzzling symptoms.

Contents

Abbreviations: AMPA receptors, α-amino-3-hydroxyl-5-methyl-4-isoxazole-propionate receptors; D_1 receptors, the D_1 subtype of dopamine receptors; D_2 receptors, the D_2 subtype of dopamine receptors; DLPFC, dorsolateral prefrontal cortex; fMRI, functional magnetic resonance imaging; NMDA receptors, N-methyl-D-aspartic acid; NAC, N-acetyl-cysteine; PFC, prefrontal cortex; OFC, orbitofrontal cortex; STS, superior temporal sulcus; TPJ, temporoparietal junction; VTA, ventral tegmental area.

* Corresponding author.
E-mail address: philip.corlett@yale.edu (P.R. Corlett).

0301-0082/$ – see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.pneurobio.2010.06.007

1. Introduction

Delusions are the extraordinary and tenacious false beliefs suffered by patients with various ailments ranging from schizophrenia (Schneider, 1959), to traumatic brain injury (Coltheart et al., 2007), Alzheimer's (Flint, 1991) and Parkinson's disease (Ravina et al., 2007), the ingestion of psychotogenic drugs (Corlett et al., 2009a) and, less frequently, autoimmune disorders such as Morvan's syndrome (Hudson et al., 2008) or potassium channel encephalopathy (Parthasarathi et al., 2006). Given this range of potential diagnoses, each with its own candidate neuropathology, it is perhaps unsurprising that we have not converged upon an agreed neurobiology of delusions. Delusions are particularly hard to study because of their insidious onset and tonic nature, their conceptual rather than behavioral basis (making them difficult to study using animal models), and the absence of a coherent theoretical model. We aim to address these issues in the current review by developing a translational model of delusion formation which we believe makes delusions tractable for animal modeling, amenable to investigation with functional neuroimaging and grounded within a theoretical framework that makes testable predictions.

Our task is made more difficult when one considers the range of odd beliefs from which people suffer; fears of persecution by clandestine forces (Melo et al., 2006); beliefs that televisions or newspapers are communicating a specific and personal message (Conrad, 1958b; Startup and Startup, 2005), the conviction that one's thoughts and movements are under the control of an external agent or are broadcast out loud (Schneider, 1959); an unrealistic belief in one's own fame or power (Karson, 1980; Kraeplin, 1902), that one is infested with parasites (Thiebierge, 1894) or deceased (Cotard, 1880), or the subject of a stranger's love (De Clerambault, 1942), or that family members have been replaced by imposters or even robots (Capgras and Reboul-Lachaux, 1923).

We take a cognitive neuropsychiatric approach to delusions. That is, the starting point is to review what we understand about the healthy functioning of a particular process, e.g. familiar face recognition, before extrapolating to the disease case, when face recognition fails and delusions of misidentification form (Halligan and David, 2001). This approach has proven successful for explaining certain delusions (Ellis and Young, 1990) but not yet for delusions in general. Perhaps this is because there are difficulties defining delusions as well as deciding what they have in common (if anything) with normal, healthy beliefs (Berrios, 1991; Delespaul and van Os, 2003; Jones, 2004; Owen et al., 2004). Beliefs are not easily accessible to the techniques of neuroscience which are more suited to representing states with clear experiential boundaries (Damasio, 2000; Knobel et al., 2008).

Furthermore, delusions are difficult to model in animals, given that they involve dysfunctions of what many consider uniquely human faculties like consciousness, language, reality monitoring and meta-cognition (Angrilli et al., 2009; Berrios, 1991; Moritz et al., 2006). Computational models of core cognitive functions (such as working memory) are being applied to gain insights into neural dysfunction in schizophrenia (Seamans and Yang, 2004; Winterer, 2006) and some are beginning to address the phenomenology of specific psychotic symptoms (Loh et al., 2007), however, these models have focused on circuit mechanisms within a local area (like prefrontal cortex), they are unable to capture the content of particular symptoms which involve information processing across large networks of interacting brain regions (Fuster, 2001).

There is a need for a testable conceptual model of delusions, one that is rooted in translational cognitive neuroscience. We, and others, propose that beliefs (both normal and abnormal) arise through a combination of innate or endowed processes, learning, experience and interaction with the world (Friston, 2010). Like other forms of information, beliefs are represented in the brain through the formation and strengthening of synaptic connections between neurons, for example causal beliefs may be mediated by a strengthening of the synaptic associations between pools of neurons representing a particular cause and their counterparts representing an associated effect (Dickinson, 2001; McLaren and Dickinson, 1990; Shanks, 2010). There are neural (and hence cognitive) limits set on the range of possible connections that can be made (Kandel, 1998). The strength of those connections is modifiable such that those conveying an adaptive advantage are strengthened and those that are disadvantageous are weakened (Hebb, 1949b; Thorndike, 1911).

This set of sculpted connections is used to predict subsequent states of the internal and external world and respond adaptively (Friston, 2005b); however, should that next state be surprising, novel or uncertain new learning is required (Schultz and Dickinson, 2000). Our premise is based upon the idea that the brain is an inference machine (Helmholtz, 1878/1971) and that delusions correspond to false inference. This inference is necessarily probabilistic and rests upon some representation of predictions (prediction error) and uncertainty (i.e. precision) about those predictions. Within this framework, we see delusions as maladaptive beliefs that misrepresent the world. They might arise through any number of perturbations within this scheme, from an unconstrained specification of the possible or lawful set of neural connections (Hoffman and Dobscha, 1989); providing the potential for bizarre beliefs to form (Hemsley and Garety, 1986a), to an adventitious and inappropriate reinforcement of particular neural connections (King et al., 1984; Shaner, 1999); engendering unexpected percepts, attentional capture and beliefs that deviate grossly from reality (Corlett et al., 2009a, 2007a; Fletcher and Frith, 2009). Impaired predictive mechanisms have been previously implicated in delusions of alien control; whereby the sufferer believes their movements are under the control of an external agent because of an inability to appropriately predict the sensory consequences of their actions (Frith et al., 2000b). We propose that this account generalizes from actions to numerous cognitive processes, that predictive learning and prediction errors are general mechanisms of brain function (Friston, 2005b; Schultz and Dickinson, 2000) and that aberrant predictions and prediction errors provide a unifying explanation for delusions with disparate contents.

A crucial distinction, which we will appeal to repeatedly, is between prediction errors per se and the precision or uncertainty about those errors. We will develop the argument that delusions (and their neurotransmitter basis) represent a failure to properly encode the precision of predictions and prediction errors; in other