compelling neurocognitive model – Perception and Attention Deficit (PAD) – with which to account for a range of their phenomenological, pathological, and clinical features. Yet the success of the model comes at an expense. For it to succeed, the authors have been forced to make it something of a Procrustean bed, stretching some parts of the visual hallucination evidence and amputating others. This is no more apparent than in their attempt to deal with the hallucinations associated with eye disease, the third-ranking pathological condition in their estimates of population-wide morbid load, exceeded only by delirium and the combined dementias.

As acknowledged by the authors, in eye disease (or, indeed, any visual-pathway lesion), simple hallucinations far outnumber complex ones, an observation which contrasts with core PAD disorders such as Lewy body dementia, Alzheimer's disease, and Parkinson's disease, in which complex hallucinations far outnumber simple ones. For eye and visual-pathway disease, complex hallucinations are only a small part of a much larger clinical picture, their importance needing to be stretched to match core PAD disorders. In fact, even after stretching, the match is an imperfect one. For example, complex hallucinations such as figures in patients with eye disease tend to be bizarre and unfamiliar, often wearing elaborate costumes and hats (Santhouse et al. 2000). In contrast, the figures hallucinated in Parkinson's disease and the dementias tend to be mundane and familiar (Fénelon et al. 2000). Yet, perhaps the most serious objection to including eye-related hallucinations in the PAD model is not the stretched importance of complex hallucinations but the amputation of their simple counterparts. All visual hallucinations, whether simple or complex, relate to phasic increases in activity within visual cortex, the difference between the two categories being the location of the activity increase. For example, activity increase in the human colour centre V4 will result in the hallucination of a "simple" formless coloured blob, whereas activity increase a few centimetres anterior to V4, in object-specialised cortex, will result in the hallucination of a "complex" object (ffytche et al. 1998). An important weakness of the PAD model is that it is forced to make an arbitrary distinction between these different cortical loci and their related hallucinations, amputating from its remit cortical areas underlying the simple hallucinations which typify those found in eye disease.

There are other features of visual hallucinations which require amputation for eye and visual pathway disease to fit the PAD model. In eye and visual-pathway disease, visual hallucinations, whether simple or complex, tend to resolve over time, with 60% of patients with visual hallucinations related to eye disease being hallucination-free at 18 months (Holroyd & Rabins 1996) and almost all patients with visual hallucinations related to visual-pathway infarcts being hallucination free within weeks (Kölmel 1985). Such patients do not develop elaborate delusional explanations for the experiences and typically gain insight into their hallucinatory nature even if, initially, some believe them to be real. These visual hallucinations are invariably silent and are not interspersed with hallucinations in other sense modalities. This overall clinical picture is sufficiently characteristic that exceptions to it point to the presence of other, non-ophthalmic causes for the hallucinations. Contrast this with the clinical picture found in the dementias, Parkinson's disease and schizophrenia. Here the visual hallucinations tend to persist or progress with time (see, for example, Goetz et al. 2001b in the context of Parkinson's disease) and are typically associated with insightless, delusional explanations. The visual hallucinations in these conditions tend to be associated with other sense modalities, either simultaneously (e.g., seeing and hearing the hallucination) or on different occasions (e.g., visual hallucinations interspersed with auditory hallucinations). Indeed, it is something of a psychiatric axiom that visual hallucinations in schizophrenia never occur without auditory hallucinations, either as separate hallucination events or as simultaneous, multimodality hallucinations. The PAD model is forced to ignore these striking clinical differences to allow the visual hallucinations of one set of conditions to sit comfortably with those of another.

Without stretching and amputation, what seems to emerge from the visual hallucination evidence taken as a whole is two distinct, overlapping syndromes.[1] One syndrome consists of predominantly simple hallucinations which resolve with time, occur with insight and without delusions, and are purely visual. The second consists of predominantly complex hallucinations which persist over time, occur with delusions and without insight, and cross sensory modalities. Setting aside those conditions in which the visual cortex is stimulated directly (e.g., migraine and epilepsy), to a first approximation all clinical conditions in which visual hallucinations occur are associated with one or other of these syndromes: eye and visual-pathway disease to the first, core PAD conditions to the second.

The existence of two distinctive syndromes poses a significant challenge to PAD and other models that treat visual hallucinations as a single pathophysiological entity. It seems unlikely that two such very different symptom profiles could emerge from the same disordered mechanism.

Perhaps the time for unitary models of visual hallucinations has passed. If there are two syndromes of visual hallucinations rather than one, we need to broaden our explanatory accounts to allow for this dichotomy. One approach would be to include two distinct but interacting pathophysiological mechanisms into our models, each related to one of the two syndromes. Obvious candidates would be visual de-afferentation as underlying the first syndrome and PAD cholinergic dysfunction the second syndrome (ffytche 2004; 2005). If correct, such expanded, bipartite pathophysiological models have important implications. In the clinic, they suggest, unlike their unitary counterparts, that different types of visual hallucinations need different treatment strategies (ffytche 2004). In the laboratory, and perhaps more significantly, by providing a comprehensive account of the neural mechanisms of disordered conscious vision, such extended models take us a step closer to a neural account of visual consciousness.

NOTES

**1.** These syndromes are unrelated to those associated with eye disease described in Santhouse et al. (2000), which would be considered sub-syndromes of the first, predominantly simple hallucination syndrome described here.

## Hallucinations and perceptual inference

Karl J. Friston

*Department of Imaging Neuroscience, Institute of Neurology, London WC1N 3BG, United Kingdom.* **k.friston@fil.ion.ucl.ac.uk**
**http://www.fil.ion.ucl.ac.uk**

**Abstract:** This commentary takes a closer look at how "constructive models of subjective perception," referred to by Collerton et al. (sect. 2), might contribute to the Perception and Attention Deficit (PAD) model. It focuses on the neuronal mechanisms that could mediate hallucinations, or false inference – in particular, the role of cholinergic systems in encoding uncertainty in the context of hierarchical Bayesian models of perceptual inference (Friston 2002b; Yu & Dayan 2002).

Collerton et al. provide a compelling synthesis implicating cholinergic dysfunction in the aetiology of recurrent complex visual hallucinations (RCVH). Furthermore, they observe "that both sensory release and top-down activation are necessary, but neither in itself is sufficient to cause high rates of RCVH" (sect. 6.3, para. 3). This fits very comfortably with models of perceptual inference based on hierarchical Bayes, in which cholinergic mechanisms may balance bottom-up sensory evidence and top-down priors by encoding their relative uncertainty or precision. In short, cholinergic dysfunction may result in a failure to properly integrate sensory information and prior expectations. In what follows, I try to explain how this might happen.

Perceptual inference is the same as statistical inference and

rests on the probability density of the causes of sensory information (i.e., the conditional probability). In classical inference, using, say, *t*-tests, inference is based on two things: (1) an estimate of the effect and (2) the standard error or uncertainty about that estimate. The *t*-statistic is simply the ratio of these two quantities. The basic idea here is that hallucinations can be regarded as false inference that arises not because of impaired estimation (i.e., sensation) but a failure to encode the uncertainty. In the *t*-test example, this might mean the standard error was always too small, leading to false inference based on pathologically large *t*-values. How might this happen in the brain?

Current thinking in computational neuroscience and machine learning points to hierarchical Bayes as the best candidate for understanding perception. I have introduced the notion of empirical Bayes in this context: empirical Bayes using the conditional independence among hierarchical levels to form empirical priors based on the sensory data. This means (almost paradoxically) that cortical hierarchies can construct their own priors, where each level of the hierarchy is subject to constraints or priors from the level above (top-down effects) when accounting for sensory evidence from below (bottom-up effects). There are many issues that attend this theoretical perspective (see Friston 2002b for review). Here I focus on the putative role of cholinergic neurotransmission in the genesis of hallucinations.

Mathematically, neuronal dynamics and synaptic efficacy are considered to minimise something called the *free energy* ($F$, a concept from statistical physics). The quantities that minimise the free energy are the conditional density $q(\nu)$ of the causes $\nu$ of sensory input (e.g., a high-level representation of a face) and some hyperparameters $\lambda$ encoding the uncertainty or noise. These two quantities correspond loosely to the numerator and denominator of the *t*-statistic above and are updated in two iterated steps: the **E**-step and the **M**-step. This is known as *expectation maximisation* in statistics.

$$\mathbf{E} \quad q(\nu) = \min_{q} F$$

$$\mathbf{M} \quad \lambda = \min_{\lambda} F$$

For a hierarchical model, the **E**- and **M**-steps for the $i$-th level can be implemented with the following descent scheme, for any generative or constructive causal model $\nu_i = g_i(\nu_{i+1})$ under Gaussian assumptions:

$$\mathbf{E} \quad \frac{\partial \hat{\nu}_i}{\partial t} = -\frac{\partial F}{\partial \nu_i} = -\frac{\partial \xi_{i-1}^T}{\partial \nu_i}\xi_{i-1} - \frac{\partial \xi_i^T}{\partial \nu_i}\xi_i$$
$$\xi_i = \hat{\nu}_i - g_i(\hat{\nu}_{i+1}) - \lambda_i \xi_i$$

$$\mathbf{M} \quad \frac{\partial \lambda_i}{\partial t} = -\frac{\partial F}{\partial \lambda_i} = -\left\langle \frac{\partial \xi_i^T}{\partial \lambda_i}\xi \right\rangle - (1+\lambda_i)^{-1}$$

This can be implemented in a simple neuronal architecture of the sort shown in Figure 1. Here the conditional density is represented in terms of its average or expectation $\hat{\nu}_i$ and covariance $\Sigma_i$, *i.e.*, $q(\nu_i) = N(\hat{\nu}_i, \Sigma_i)$ where

$$\sum_i^{-1} = \frac{\partial \xi_i^T}{\partial \nu_i}\frac{\partial \xi_i}{\partial \nu_i} + (1+\lambda_i)^2 I$$

which is an implicit function of the hyperparameters. In this scheme, the quantities $\hat{\nu}_i$ and prediction error $\zeta_i$ correspond to the activity of two neuronal subpopulations, whereas the hyperparameters $\lambda_i$ are encoded by the synaptic efficacy of lateral connections.[1] Note that this scheme converges when $\hat{\nu}_i$ cannot further reduce prediction error and $\partial \zeta_i / \partial \nu_i^T \zeta_i = 0$. In Friston (2002b) I discuss the potential role of cholinergic neurotransmission in mediating the **M**-step. A related theme, using a different perspective, is discussed in Yu and Dayan (2002). What would happen if the hyperparameters were encoded improperly with cholinergic dysfunction?

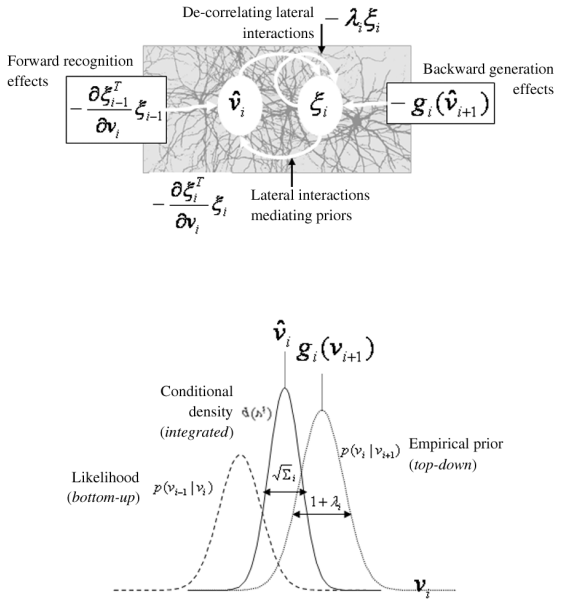


Figure 1 (Friston). The top panel is a schematic showing two neuronal subpopulations representing the conditional expectation of sensory causes for a single cortical level and the influences they are subject to. The bottom panel shows the implicit probability densities encoded by these neuronal activities and synaptic efficacies after convergence. Note that the uncertainty or width of these densities is determined by the hyperparameters. The conditional density, upon which inference is based, is drawn in a solid line.

A failure to optimise the hyperparameters will produce an inappropriate balance between sensory and prior influences on the conditional expectation of what caused any sensation. This is shown schematically in Figure 2. Here, we assume the deficit produces hyperparameters that fail to encode uncertainty in the priors. This means too much weight is afforded to the prior expectation from supraordinate cortical levels, and false inference ensues. Collerton et al. discuss a similar notion from the point of view of a "failure to select the correct proto-object in the PAD model"

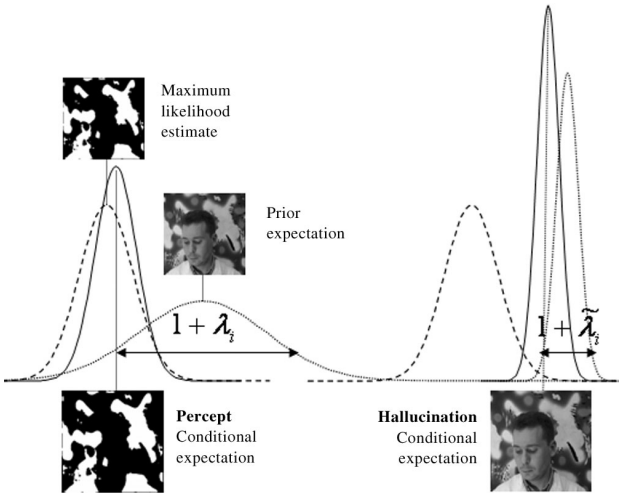


Figure 2 (Friston). A schematic showing one way in which hallucinations could occur. In this example, the hyperparameter encoding prior uncertainty has been made too small $\lambda_i \rightarrow \tilde{\lambda}_i$, resulting in overconfidence in the priors and a false or hallucinatory conditional expectation.

(sect. 7.4.2, para. 2) when cholinergic inhibition leads to incorrect "pattern matching."

The mechanistic understanding afforded by this computational approach can usefully account for many observations made by Collerton et al. For example, "Either impaired attention [i.e., prior expectations] or impaired sensory activation [i.e., evidence] alone will rarely produce hallucinations" (sect. 7.3, para. 1). It is their relationship that defines a hallucination. In this sense, the integration, through the conditional density, is the key mechanism in perception and this integration may depend on the integrity of cholinergic mechanisms. The false learning associated with more enduring changes mediated by the **M**-step may improperly pair sensory contexts with high-level representations leading to "the same image being triggered again and may account for the repetition of specific images" (sect. 7.5.1, last para.). In empirical Bayes the priors are driven by prediction errors from the level below (see Fig. 1). In the absence of sensory input, priors are not induced. This may account for what the target article describes as "an otherwise puzzling feature of hallucinations – that they disappear on eye closure or on complete visual loss" (sect. 7.5.2, para. 2).

In terms of clinical neuroscience, there are remarkable overlaps between the PAD model for hallucinations and the disconnection hypothesis for schizophrenia, a disorder associated with hallucinations. In terms of functional anatomy, Collerton et al. note that "Object-based attention depends primarily on the function of lateral frontal cortex, and object perception depends primarily on the ventral visual stream" (sect. 7.3, point 3). They later cite evidence from functional imaging of patients who are prone to hallucinations. In fact, the disconnection hypothesis was based on early observations of abnormal coupling between left dorsolateral prefrontal cortex and posterior temporal regions, as measured with positron emission tomography in schizophrenics (see Friston 1998 for review).

The disconnection hypothesis posits abnormal functional integration (at the synaptic level) as the primary pathophysiological mechanism in schizophrenia. The premise is that synaptic plasticity is regulated abnormally during emotional and perceptual learning. The abnormal regulation probably involves dopaminergic dysfunction in emotional learning or operant conditioning (i.e., the formation of stimulus-response links) and cholinergic dysfunction in perceptual learning (i.e., the formation of stimulus-stimulus associations). Exactly the same neurotransmitters are implicated by Collerton et al. in RCVH: "pharmacological data so far available indicate a primary role for cholinergic and secondary role for dopaminergic dysfunction in the aetiology of RCVH" (sect. 3.2, last para.). However, they later note "that dopamine receptors are not prevalent in visual processing areas (whereas muscarinic cholinergic receptors are)" (sect. 7.4.2, para. 2). This is consistent with the conclusion of a recent editorial on disconnection and cognitive dysmetria in schizophrenia: "In short, normal interactions between dopamine and the cellular or synaptic mechanisms responsible for plasticity are essential for emotional learning, whereas the interaction between cholinergic neurotransmission and associative plasticity is important for perceptual learning" (Friston 2005). Although Collerton et al. state, "Eye disease and schizophrenia pose greater challenges to our model" (sect. 7.6.2.2, para. 5), there are encouraging and important points of contact between the PAD model and theoretical treatments of cerebral pathology in schizophrenia.

ACKNOWLEDGMENT

The Wellcome Trust funded this work.

NOTE

**1.** In this summary I have assumed that the parameters of the generative model of how sensory inputs are caused have already been learned (in the **M**-step). These parameters are encoded by the synaptic efficacy of forward and backward connections linking levels.

## Waking hallucinations could correspond to a mild form of dreaming sleep stage hallucinatory activity

Claude Gottesmann

*Département de Biologie, Université de Nice-Sophia Antipolis, 06108 Nice Cedex 2, France.* **gottesma@unice.fr**

**Abstract:** There are strong resemblances between the neurobiological characteristics of hallucinations occurring in the particular case of schizophrenia and the hallucinatory activity observed during the rapid-eye-movement (dreaming) sleep stage: the same prefrontal dorsolateral deactivation; forebrain disconnectivity and disinhibition; sensory deprivation; and acetylcholine, monoamine, and glutamate modifications.

To explain the neurobiological deficiencies responsible for hallucinations, the PAD model described by Collerton et al. first highlights attention impairments. These could be related to a prefrontal dorsolateral greater or lesser deactivation also observed in schizophrenia (Bunney & Bunney 2000; Lewis 2000; Weinberger et al. 1986). An analogy can be made with the rapid-eye-movement (REM) dreaming sleep stage, a possible model of schizophrenia (Gottesmann 2002; 2004a; 2004b). During this sleep stage, hallucinatory activity also occurs, as evidenced in cats (Henley & Morrison 1974; Jouvet & Delorme 1965; Jouvet & Mounier 1960; Sastre & Jouvet 1979) and rats (Mirmiran 1983; Mouret & Delorme 1967; Sanford et al. 2001) after experimental suppression of usual muscular atonia, and in normal subjects (Aserinsky & Kleitman 1953) as well as, similarly, in the pathological so-called REM Sleep Behavior Disorder (Mahowald & Schenck 2004), which is the human form of REM sleep without atonia. Indeed, a specific inactivation of the same dorsolateral prefrontal cortex (Braun et al. 1997; Maquet et al. 1996) is observed in this sleep stage. Moreover, whether a cause or a consequence, and during REM sleep as opposed to waking, the frontal cortex seems to be disconnected from other cortical areas, particularly perceptual areas, as the gamma rhythm becomes uncoupled over cortex areas (Perez-Garci et al. 2001). Here, also, there is a strong resemblance with processes occurring in schizophrenia, given that intracerebral disconnections have long been hypothesized to explain the symptoms in this mental illness (Meyer-Lindenberg et al. 2001; Peled et al. 2000; Young et al. 1998).

Therefore, it is of interest that the PAD model also associates this prefrontal cognitive impairment with a decrease in perception processes. This symptom was already described for hallucination development by several authors. More particularly, Behrendt and Young (2004) recently reported a thalamus unconstrained by the usual sensory afferents. Here again, a parallel can be drawn with REM sleep. Indeed, Dement (1958) first identified the increase of arousal threshold by peripheral stimuli, which indicates that this sleep stage corresponds to deep sleep; this was also shown by the difficulty of arousal after central stimulation (Benoit & Bloch 1960). However, more precise experimental arguments have strengthened the notion of perception deficit underlying hallucinations. The sensory de-afferentation hypothesis is strongly supported by the presynaptic inhibition observed during the REM sleep stage in the thalamic relay nuclei of cats (Steriade 1970) and rats (Gandolfo et al. 1980). This failure of sensory afferents during REM sleep is further reinforced at the cortical level. Indeed, while the associative visual areas that lead to the ventral visual stream involved in the PAD model are activated during REM sleep, the primary visual cortex, the target of sensory afferents, is deactivated (Braun et al. 1998).

The neurochemical model of the PAD suggests that hallucination occurrence is based on a decrease of acetylcholine and an excess of dopamine functioning. In our REM sleep neurobiological model of schizophrenia, it has to be emphasized that the cortical release of acetylcholine is lower than during active waking (Marrosu et al. 1995), which could explain a cognitive impairment, although the acetylcholine level is higher in the basal forebrain