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 $q(v)$ of the causes v of sensory input (e.g., a high-level representation of a face) and some hyperparameters λ 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.

$$\mathbf{E} \quad q(v) = \min_q F$$

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

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 $v_i = g_i(v_{i+1})$ under Gaussian assumptions:

$$\mathbf{E} \quad \frac{\partial \hat{v}_i}{\partial t} = -\frac{\partial F}{\partial v_i} = -\frac{\partial \xi_{i-1}^T}{\partial v_i}\xi_{i-1} - \frac{\partial \xi_i^T}{\partial v_i}\xi_i$$

$$\xi_i = \hat{v}_i - g_i(\hat{v}_{i+1}) - \lambda_i\xi_i$$

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

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 $\hat{v}_i$ and covariance Σ_i, i.e., $q(v_i) = N(\hat{v}_i, \Sigma_i)$ where

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

which is an implicit function of the hyperparameters. In this scheme, the quantities $\hat{v}_i$ and prediction error ζ_i correspond to the activity of two neuronal subpopulations, whereas the hyperparameters λ_i are encoded by the synaptic efficacy of lateral connections.[1] Note that this scheme converges when $\hat{v}_i$ cannot further reduce prediction error and $\partial \zeta_i/\partial v_i^T \zeta_i = 0$. 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?
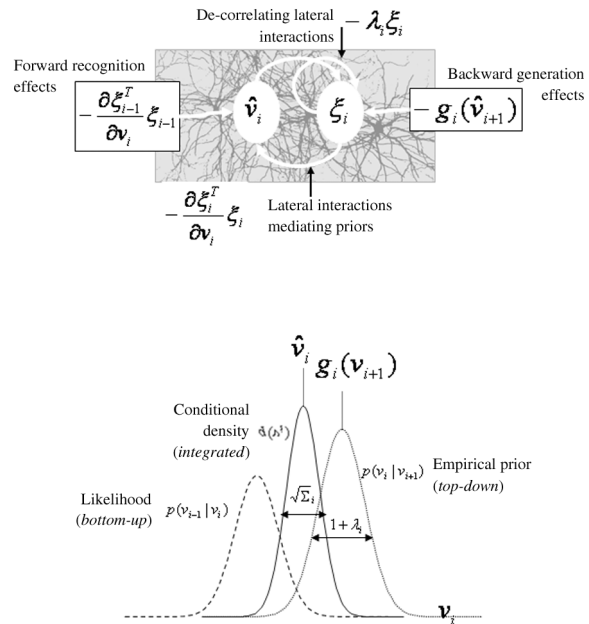

Figure 1 (Friston). The top panel is a schematic showing two neuronal subpopulations representing the conditional expectation of sensory causes for a single cortical level and the influences they are subject to. The bottom panel shows the implicit probability densities encoded by these neuronal activities 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"
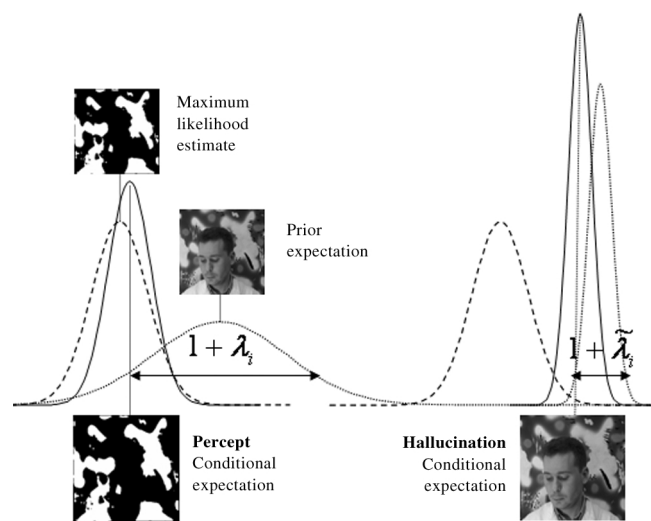

Figure 2 (Friston). A schematic showing one way in which hallucinations could occur. In this example, the hyperparameter encoding prior uncertainty has been made too small $\lambda_i \to \tilde{\lambda}_i$, resulting in overconfidence in the priors and a false or hallucinatory conditional expectation.

(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.

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.