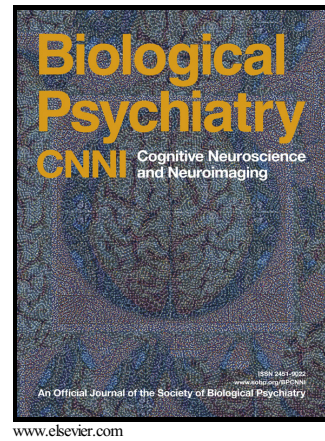


Author's Accepted Manuscript

Hallucinations as Top-Down Effects on Perception
Top-down Hallucinations?

R. Powers Albert, Megan Kelley, Philip.R. Corlett



PII: S2451-9022(16)30031-3
DOI: <http://dx.doi.org/10.1016/j.bpsc.2016.04.003>
Reference: BPSC50

To appear in: *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*

Received date: 1 February 2016
Revised date: 22 April 2016
Accepted date: 22 April 2016

Cite this article as: R. Powers Albert, Megan Kelley and Philip.R. Corlett, Hallucinations as Top-Down Effects on Perception *Top-down Hallucinations?*, *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, <http://dx.doi.org/10.1016/j.bpsc.2016.04.003>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting galley proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Title: Hallucinations as top-down effects on perception

Running title: *Top-down Hallucinations?*

Powers III, Albert, R.¹; Kelley, Megan¹; Corlett, Philip.R.*¹

*: For correspondence: Philip Corlett, Ph.D., Abraham Ribicoff Research Facilities, Connecticut Mental Health Center, 34 Park St, New Haven, CT 06511; (tel) 1-203-974-7866; email: Philip.corlett@yale.edu

Author affiliations:

1. Department of Psychiatry, Yale University

Keywords:

Psychosis

Hallucinations

Predictive Coding

Neuroimaging

Connectivity

Modularity

Neuropsychology

ABSTRACT: The problem of whether and how information is integrated across hierarchical brain networks embodies a fundamental tension in contemporary cognitive neuroscience, and by extension, cognitive neuropsychiatry. Indeed, the penetrability of perceptual processes in a ‘top-down’ manner by higher-level cognition—a natural extension of hierarchical models of perception—may contradict a strictly modular view of mental organization. Furthermore, some in the cognitive science community have challenged cognitive penetration as an unlikely, if not impossible, process. We review the evidence for and against top-down influences in perception, informed by a predictive coding model of perception and drawing heavily upon the literature of computational neuroimaging. We extend these findings to propose a way in which these processes may be altered in mental illness. We propose that hallucinations - perceptions without stimulus - can be understood as top-down effects on perception, mediated by inappropriate perceptual priors.

Introduction

Imagine you are walking home on a warm early summer night. The sights and sounds that greet you are familiar—the bark of your neighbor’s dog, the old oak tree on the corner, the echo of your footsteps as you get closer to your destination. Now imagine you are walking the same route after watching a scary movie. The same things might now seem strange and menacing to you. The dog’s bark might seem like a growl; the oak tree’s shadows may seem more prominent; those echoed footsteps might sound louder—or maybe they might not seem to be your own. Your aroused state makes you search for hidden threats, and your beliefs guide where you search (1). Could your fear even lead you to hallucinate footsteps when there are none? We consider these questions of how cognition alters perception, in light of recent advances in computational psychiatry.

Present-day cognitive scientists have argued that cognition does not influence perception (2). However, work in computational neuroscience calls this claim into question. We will argue that hallucinations too challenge strict, encapsulated modularity. Instead, they indicate penetrability of perception by cognition. We will illustrate these claims with phenomenology as well as neuro-computational work.

Modules and the Mind

In *The Modularity of Mind* (1983), Jerry Fodor sketched a blueprint of mental architecture comprised of modules—systems that process a single specific kind of information (3). For example, the early vision module takes in ambient light and outputs color representations. Fodor never gave a specific definition of a module (nor have other modularists), which makes the theory difficult to falsify. Twyman and Newcombe write: “*Given this lack of agreed-upon definition, the modularity position becomes analogous to the Hydra, the many-headed monster that Heracles found difficult to combat because there were too many heads to take on simultaneously, and, worse, because other heads grew while he addressed a specific one*” (4). Ultra-cognitive neuropsychologists even claim to shun brain localization, calling it mere hardware and irrelevant to the software they are interested in (5). However, even the most ardent ultra-cognitivists utilize some lesion location information in rendering their arguments (6).

A strictly modular approach focuses on functional segregation, with brain regions responsible for discrete mental faculties that can be damaged in isolation (7). Such an approach eschews functional integration, which posits that complex mental functions are based on interactions among distributed regions (7). Human lesion cases also support integration (8-10). We appeal to functional and effective connectivity data for insights into integration. We view this work via a model of mind and brain that itself challenges encapsulated modularity—namely, predictive coding (11).

Predictive Perception Implies Cognitive Penetration

An encapsulated perceptual system, kept separate from the influence of beliefs, could have the advantage of keeping our beliefs grounded in the truth offered by our senses (12). However, a cognitively penetrable perceptual apparatus may be equally adaptive, despite misperceiving and misbelieving, as long as the resulting behavior is adaptive (13, 14). Hume (15) and Helmholtz (16) appreciated this: we perceive what would need to be present in order for our sensations to make sense. To solve this inverse problem the brain uses an internal generative model of its environment to infer what it is sensing (16), combining feed-forward “bottom-up” information from sensory organs with feedback or “top-down” predictions from higher-level regions to compute precision-weighted prediction errors that guide formation of an optimal estimation of the surroundings (17-20). Combining top-down expectation and bottom-up input to explain perception has a rich history. McClelland and Rumelhart proposed early models with this motif (21, 22). Rao and Ballard added neural data and Bayes theorem (23). Maia and Cleeremans proposed that perception solves a ‘global constraint satisfaction’ problem via the interplay between current top-down prefrontal cortical modulation and prior knowledge through learned synaptic connections across a hierarchy (24). Friston first described how these mechanisms may be generalized to a broad model of brain function in terms of predictive coding (25).

Contrary to encapsulated modularity, some studies claim that early visual processing (i.e., not “post-perceptual processing”) is influenced by non-perceptual information (27-30). For example, semantic priming increases the speed and accuracy of detection by minimizing prediction error (28). Behavioral and neurophysiological evidence shows prediction error signals generated in early visual regions in response to violations of semantic expectation (31, 32). Word contexts result in ambiguous shapes being perceived as the missing letters that complete a word (33, 34). Semantic categories including letters and animals improve accuracy and response-times in orientation identification (28, 35). Audiovisual integration induces response changes in primary sensory cortices, such that auditory stimuli engage V1 and visual stimuli activate A1 (36). These activations evolve via prediction error-driven learning (36). These phenomena challenge the informational encapsulation of perception (11).

The Burden of Proof: Establishing Top-Down Influences in Perception

Studies that comprise the so-called *New Look* movement, purporting to demonstrate effects of language and culture on perception, have recently come under scrutiny. Firestone and Scholl (2) asserted such studies may be plagued by confounds that can be avoided by following these guidelines: 1) Disentangle perceptual from decisional processes; 2) Dissociate reaction time effects from primary perceptual changes; 3) Avoid demand characteristics 4) Ensure adequate low-level stimulus control; and 5) Guarantee equal attentional allocation across conditions.

These suggestions address issues inherent to tasks where perception guides a behavioral decision. However, Bayesian formulations do not accept such separation (37). Signal detection theory appears to sharply distinguish sensation from decision. However, it too allows cognition to influence perception (38). Top-down processes can even alter the mechanical properties of sensory organs (39) by altering the signal-to-noise ratio (40). As we will argue in the following sections, top-down influences may be clearest when sensory input is completely absent: when experiences are hallucinated.

Hallucinations as Examples of Top-Down Penetration

Hallucinations (41) can have contents consistent with one's affective state (42). When people are depressed, hallucinations can contain themes like guilt and disease. Those experienced during mania center around extraordinary powers (43). Changes in the content of hallucinations can be wrought by experimental mood manipulations (44). Thus affect may penetrate perception. However, auditory-verbal hallucinations (AVH) represent a derangement of normal function. Perhaps perception is normally impenetrable.

The existence of “non-clinical voice-hearers”—who have auditory hallucinations but do not reach clinical criteria for a psychotic disorder—argue against this hypothesis. Hallucinations of a loved one are common following bereavement (45-47). They are typically comforting and do not impair functioning (45, 46) (although see (47)); thus, hallucinations may not be abnormal *per se*. Non-clinical hallucinations also occur in the general population (48-52). Estimates of their prevalence are as high as 28% (53) and only 25% of those meet the diagnostic criteria for a psychotic disorder (54). Thus, hallucinations may best be described as an extreme of normal functioning (48) rather than a failure of modularity.

Are hallucinations top-down processes? In a recent investigation (55), prior knowledge of a visual scene conferred an advantage in recognizing a degraded version of that image. Patients at risk for psychosis were particularly susceptible to this advantage. A bias toward top-down information is the basis for “sensory conditioning” (56-60), wherein a visual stimulus is established as a predictor of a difficult-to-detect auditory stimulus and participants begin to report auditory stimuli that were not presented on the basis of the visual cue. This effect is amplified in individuals who hallucinate (57). Experiences of uncertainty can increase top-down effects. When a participant's sense of control over the environment is intentionally decreased (with spurious feedback), they tend toward illusory pattern perception, seeing nonexistent signal in noise and detecting illusory trends in the stock market (61).

The guidelines proffered by Firestone and Scholl (2) may serve as a useful roadmap for future studies of perceptual decision-making tasks. However, studying the penetrability of perception by way of hallucinatory experiences may circumvent these pitfalls. Participants report spontaneous, vivid experiences rich with sensory information that are unlikely to result from attentional biases. Neuroimaging data may likewise circumvent some critiques. We now try to integrate our understanding of hallucinations with notions of neural modularity and connectivity.

Brain lesions, modularity, connectivity and hallucinations.

We propose that inter-regional effects mediate the penetration of perception by cognition. Some have discussed these top-down effects in terms of attention (2). Predictive coding theory conceives of attention in terms of the precision of priors and of prediction errors. Formally this could be implemented with a Kalman Filter (62, 63), which combines the predictive relationship between states (reliability), together with how this relationship is expected to change over time (the uncertainty) (64). This formalism can explain behavioral neuroscience observations in predictive learning (64). There are also encouraging human data. (65). However, precision weighting awaits more extensive empirical investigation in humans (particularly with regards to neurobiology and neurochemistry).

The precision of priors, in our example from the Introduction, may guide the search for threats in the shadows after having watched a scary movie. We contend that extremely precise (i.e., strongly weighted) priors could result in perception of particularly dark shadows, and—if priors are weighted strongly enough—could result in a hallucination. The supportive data are threefold: (i) a single case where a lesion caused hallucinations; (ii) functional connectivity between the lesion location and other regions; and (iii) effective (directional) connectivity in patients who suffer AVH. With these data, we will argue that top-down priors influence perception, thus violating a strictly encapsulated modular architecture.

(i) Lesion-Induced Hallucinosi

In graph-theory approaches to network analysis, hubs are defined from network structure based on structural, functional or effective connectivity data. fMRI data analyzed using graph theoretic metrics parses the brain into hubs (sub-networks), where a subset of regions connects those sub-networks (“connectors” in **Figure 1**, above). A module in graph theory has more within-group than between-group connections. A rich club is a collection of high degree hubs, more highly interconnected than predicted by chance. Lesions that impair cognition are more likely to be in *rich-club hubs*, or regions that mediate the long-range connectivity between connected information processing hubs (66). One such rich-club hub, the limbic system (including anterior insula cortex), has recently been implicated in the global specification of precision weighting or gain control (67) and has consistently been implicated in symptom capture fMRI studies of AVH (68, 69). It would be a challenge to call these circuits part of early perception. Rather, we will show regions like orbitofrontal cortex penetrate perceptual processing in primary sensory cortices giving rise to hallucinations.

Geddes and colleagues describe visual peduncular hallucinosis in a 66-year-old man following an ischemic stroke in the left pons, encompassing locus coeruleus (LC) (70). They found that functional connectivity between the LC and visual association cortex was reduced in the hallucinating patient. Conversely, functional connectivity between brainstem and visual association cortex, and between visual association cortex and prefrontal cortex (PFC), was significantly increased. In rodents, LC sends inhibitory noradrenergic inputs to visual cortex that improve perceptual acuity by gating target neurons (71). Norepinephrine may mediate the neuronal gain-control that implements this gating (72). Gain-control involves the weighting of conflicting cortical states to arbitrate which state governs experience (73). With higher gain, neuronal responses become more feature-selective and learning rate increases (74). This would be a candidate mechanism for our search for threatening figures on our walk home: if gain is driven up too high we may perceive villains where there are none.

The hallucinating patient had hyper-connectivity between visual association cortex and orbitofrontal cortex, a

region that specifies top-down priors (75). PFC stimulation produces (76)—and frontal leucotomy ameliorates—hallucinations (77). These regions are outside of early perceptual modules.

(ii) Lesion effects on graph-theory metrics

Cognitive neuropsychology depends on the assumption that brain damage predominantly affects only the function of the damaged region. However, focal damage can disrupt whole-brain network organization (78). Intriguingly, damage to between-module connections (so-called *rich club hubs*, characterized as especially densely connected nodes (79) – see **Figure 1**) has more profound effects on cognitive performance (78) and, crucially, lesions alter connectivity in the opposing, un-lesioned hemisphere (78). The latter finding is difficult to reconcile with cognitive neuropsychology, which depends on very circumscribed damage effects (7).

We suggest that the rich-club hubs that alter global network function (and renounce encapsulation) are also the hubs involved in specifying global precision and therefore updating of inference in predictive coding (67). Patients with schizophrenia do have perturbed connectivity in rich club hubs in regions of the Default Mode Network and salience network (80), which themselves have been linked to predictive coding (81). However, hallucinations have not been specifically related to these metrics (82), and this will be an important direction for future inquiry. Rich club hubs seem well placed to mediate the top-down influence of one module on another (83). Dehaene and Changeux embrace this idea in their global workspace model of conscious cognitive functions, in which connector hub brain regions mediate informational integration into a coherent whole (84). Dysfunction of this process has been extended to understand some of the problems in schizophrenia (although not hallucinations per se (85)).

In our predictive coding approach informational integration (between modules) is mediated via precision weighting of priors and prediction errors, perhaps through rich club hubs. Others before us have asserted congruence between neural and mental modules (88). However, the exact relationship between psychological

modularity (86) and modularity in functional connectivity (87) remains an open empirical question.

(iii) Directional effects.

Thus far, the lesion effects we have discussed do not say anything about directionality. Dynamic causal modeling (DCM) is a means of making inferences about directional or effective connectivity of fMRI data as a function of extrinsic inputs (e.g., tasks) or intrinsic network states (89). Curcic-Blake and colleagues (90) examined inner speech processing in patients with schizophrenia with and without auditory hallucinations using DCM. They found a relative dearth of connectivity from Wernicke's to Broca's areas in patients with auditory hallucinations, suggesting that the precision of processing in Broca's was higher than in Wernicke's, consistent with a relative reliance on top-down versus bottom-up signals in those who were hallucinations. These data suggest information from higher regions is penetrating lower region. On a finer scale, single unit recording data have demonstrated rich-club architecture between groups of cells and that rich-clubs are key mediators of transfer entropy (directional information flow) between other hubs (83).

Chanes and Barrett (83) observe that predictions flow from less to more laminated cortices (top-down) and prediction errors flow in opposite direction (bottom-up) (91). Limbic regions like the insula have the simplest laminar structure and as such they represent good candidates for the specification of priors.

Insula may be a rich club hub (92), given its membership in a number of intrinsic circuits it is well placed to mediate the communication between intrinsic networks in the brain (92). Eldar, Cohen and Niv showed that perceptual gain (as measured by pupilometry) was correlated with clustering coefficients in functional connectivity of fMRI data (74). Correlations were particularly strong in regions with high clustering coefficients; voxels that had higher inter-correlation with other voxels (like rich club hubs). Taking these findings together, we speculate that rich club hubs are by virtue of their connectivity and functional activation, well placed to implement changes in gain control as a function of the precision of predictions and prediction errors.

Studies of bi-stable perception – percepts that switch in dominance, apparently under the influence of intrinsic factors (since nothing changes in the sensory input (93)), also support this notion. Pupil dilation predicts switches in bi-stable perception (both auditory and visual) (94). Bi-stable percepts switch more frequently in manic patients (95) and patients with schizophrenia have less volitional control over the switches between bi-stable percepts (96). With regards binocular rivalry, Tononi and Edelman made similar observations in patients with schizophrenia (97), linking their observation to alterations in functional integration and segregation in the nervous system (98). Thus far, neither changes in bi-stability nor rivalry correlate with hallucinations. This will be a key empirical focus in future work.

DISCUSSION & FUTURE DIRECTIONS

We have argued that hallucinations demonstrate the top-down penetration of perception by cognitive processes. We presented behavioral and lesion data as well as functional and effective connectivity neuroimaging data that support our thesis. These data seem incommensurate with an encapsulated modularity of mind.

We take seriously the criticisms and failed replications of *New Look* experiments (2). However, we argue that the data we raise here cannot be dismissed in terms of demand characteristics or changes to post-perceptual judgment. Furthermore, we support our case with neural connectivity suggesting that those who hallucinate have enhanced directional top-down connectivity that modulate gain control mechanisms in sensory cortices. These mechanisms go beyond enhancement of selected inputs. They may sculpt percepts in the absence of sensation.

We framed these observations in terms of predictive coding, whose broad computational goal is to minimize prediction error (25). Predictive perception is penetrable to the extent that penetration minimizes overall long-term prediction error (99), and what we know will change what we see (99): that I know the lines in the Muller-

Lyer illusion are identical does not render my perception veridical, because in the long term, the illusion is Bayes optimal (99), and knowing in this case does not minimize prediction error.

At odds with our theory, there are empirical data that demonstrate a reduced sensitivity to some illusions in patients with schizophrenia (100, 101). On the other hand, there is evidence that the influence of top-down priors is enhanced in psychotic states (102). These seem to be contradictory observations. Neither behavioral effect correlated with hallucinations specifically (weakened hollow-mask illusion correlates with disorganization symptoms, release from contextual suppression correlates with schizotypy, and enhanced impact of prior experience with stimuli correlates with schizotypal symptoms (102)).

Furthermore, other illusions may be enhanced in people with schizophrenia (for example the three flash illusion – in a manner that correlates with positive symptoms (103)). Clearly, a simple explanation of psychosis in terms of weak priors does not apply and illusions don't consistently fail in these patients. Likewise, our proposal challenges theories that posit a failure of prediction or corollary discharge in generating hallucinations. These theories have not fared well when tested empirically. Patients with schizophrenia have impaired corollary discharge, but that effect is not related specifically to hallucinations (104).

Again, hierarchy may be crucial here (105). Whilst we eschew a strict informational separation between perception and cognition, we appreciate that perceptual systems are hierarchically arranged and that information processing can be impaired at different levels of the hierarchy. Even staunch modularists allow for priors to influence perception within the perceptual module and argue that these processes mediate illusions (86). It is possible that illusions could fail at this level and hallucinations could be generated higher in the hierarchy.

Our own work with ketamine may also adjudicate. Ketamine does not usually engender hallucinations, but rather, illusions (alterations of stimuli that are present). We recently reported actual hallucinations (of music and voices) on ketamine (106). These hallucinations occurred in the MRI scanner, which is perceptually denuded (dark, still, rhythmically noisy). In prior work, we suggested (based on our functional imaging data) that ketamine enhances bottom up noise (107). In the same work, we argued sensory deprivation induces hallucinations, via top-down priors. This is similar to the paradoxical effect of hearing loss and vision loss on hallucinations. In response to low-level prediction errors (108) higher-level precision increases to accommodate, producing non-veridical percepts. The isolation of the scanner provides an ambient sensory context that interacts with the bottom-up noise under ketamine, culminating in hallucinations.

Taken together, we argue that the dynamic interaction between priors and prediction errors give rise to hallucinations (109). This dynamic interaction is weighted differently at different levels of different hierarchies. Low-level impairments may give rise to illusions (and their failure in patients with schizophrenia) and higher-level perturbations may portend hallucinations. A brain hungry for priors (102) at the lowest levels of sensory input may try to satisfy that hunger by imposing more precise priors higher in the hierarchy—weighting perception towards expectation (rather than input) in a “listening attitude,” as Arieti put it (110)—which produces hallucinations (111). Critically, these effects high and low in the hierarchy, could be somewhat separated. Similar arguments have been put forth by Deneve & Jardri (112).

Recent apologists for Fodorian modularity do allow for attention to function as a back door through which some top-down effects can occur (2). In the predictive coding framework, attention involves the precision-weighting of predictions and prediction error (17). These gain-control mechanisms may be mediated by LC noradrenergic signaling as well as corticopetal cholinergic signals, which have also been implicated in hallucinatory phenomena (113). On the other hand, GABA and glutamate signaling may be involved in divisive

normalization processes within a region or set of regions (114); this processing gates the impact of particular signals (114). Thus, within-module gain control is also possible. Our contention is that over-engagement of these processes renders priors overly precise, creating percepts in the absence of stimulus. Modulating noradrenergic, cholinergic, glutamatergic and GABAergic function may, therefore, have antipsychotic effects (115).

One criticism of cognitive penetration is that it is not under volitional control: we cannot conjure villains before our eyes at will. Some of the earliest phenomenological descriptions of schizophrenia actually describe such conjuring (of large insects for example) (116). A more recent example is found in the rise in popularity of the Tibetan spiritual practice of creating tulpas—a created companion which, through meditative practice, the creator is able to actually perceive (117). So-called ‘tulpamancy’ appears to involve a volitional choice to perceive specific agents (117). Whether tulpamancers are to be believed and whether objective evidence can be provided will require further work. By using functional imaging data and multi-voxel-pattern analysis (MVPA) to train classifiers, it may be possible to discern the neural correlates of perception from imagination (118). If tulpamancy and, more broadly, hallucinations involve truly perceptual rather than imaginative or malingering processes, they ought to engage perceptual neural correlates and be classified as such.

Further, hallucinations often involve a lack of insight although in Charles-Bonnet Syndrome they are often recognized as lacking reality (119). Seth and others have argued that the sense of perceptual presence entails not only sensory but also motor predictions (120). Even in mice, motor cortical representations modulate gain control in primary sensory cortices (121). If a percept has affordances—signifiers for action—and if those signifiers interact appropriately with our sensory feedback, then we perceive it as present (120). In this way, some hallucinations have a presence whereas others do not (119). Comparing and contrasting these phenomena is a key direction for future work. Furthermore, behavioral tasks could be devised that explore the impact of tulpas and hallucinations on behavior – rather like the studies of synesthetes that rendered them a population who may illuminate how perception is organized (122, 123).

CONCLUSION

Using functional imaging data that focus on inter-regional connectivity, a computational perspective, and focusing on hallucinations, we have made the case that perception is not encapsulated. Rather, it can be penetrated by expectations, beliefs and emotions. While this may seem an academic exercise, a deeper understanding of the neural and cognitive architecture is essential for the more precise treatment of symptoms that are distressing and portend poor functional outcome.

Acknowledgements: The authors thank Chaz Firestone, Brian Scholl, Karl Friston, Jennifer Foss-Feig, Emily Finn, Larry Marks and Emily Powers for stimulating discussions. Our acknowledgement does not imply theoretical agreement with the thesis. Any errors that remain are the authors'.

Declaration of interest: This work was supported by the Connecticut Mental Health Center (CMHC) and Connecticut State Department of Mental Health and Addiction Services (DMHAS). PRC was funded by an IMHRO / Janssen Rising Star Translational Research Award and CTSA Grant Number UL1 TR000142 from the National Center for Research Resources (NCRR) and the National Center for Advancing Translational Science (NCATS), components of the National Institutes of Health (NIH), and NIH roadmap for Medical Research. MK is part of the Yale PREP program. AP III was supported by NIMH Grant #- 5T32MH019961, "Clinical Neuroscience Research Training in Psychiatry," and by a NARSAD Young Investigator Grant. The contents of this work are solely the responsibility of the authors and do not necessarily represent the official view of NIH or the CMHC/DMHAS.

Conflicts of Interest: The authors report no biomedical financial interests or potential conflicts of interest.

REFERENCES

1. Jenkin Z, Siegel S (2015): Cognitive penetrability: Modularity, epistemology, and ethics. *Review of Philosophy and Psychology*. .6:pp.
2. Firestone C, Scholl BJ (2015): Cognition does not affect perception: Evaluating the evidence for 'top-down' effects. *The Behavioral and brain sciences*.1-77.
3. Fodor JA (1983): *The modularity of mind : an essay on faculty psychology*. Cambridge, Mass.: MIT Press.
4. Twyman AD, Newcombe NS (2010): Five reasons to doubt the existence of a geometric module. *Cogn Sci*. 34:1315-1356.
5. Coltheart M (2004): Brain imaging, connectionism, and cognitive neuropsychology. *Cogn Neuropsychol*. 21:21-25.
6. Coltheart M, Davies M (2000): *Pathologies of belief*. Oxford: Blackwell.
7. Fotopoulou A (2014): Time to get rid of the 'Modular' in neuropsychology: a unified theory of anosognosia as aberrant predictive coding. *Journal of neuropsychology*. 8:1-19.
8. Catani M, ffytche DH (2005): The rises and falls of disconnection syndromes. *Brain : a journal of neurology*. 128:2224-2239.
9. Seghier ML, Zeidman P, Neufeld NH, Leff AP, Price CJ (2010): Identifying abnormal connectivity in patients using dynamic causal modeling of fMRI responses. *Front Syst Neurosci*. 4.
10. Ween JE (2008): Functional imaging of stroke recovery: an ecological review from a neural network perspective with an emphasis on motor systems. *J Neuroimaging*. 18:227-236.
11. Lupyan G (2015): Cognitive penetrability of perception in the age of prediction: Predictive systems are penetrable systems. First Posting Jul 4, 2015. *Review of Philosophy and Psychology Jul*.
12. Quine WV, Quine WV (1951): Two Dogmas of Empiricism. *Philosophical Review*. 60.
13. McKay RT, Dennett DC (2009): The evolution of misbelief. *The Behavioral and brain sciences*. 32:493-510; discussion 510-461.
14. Johnson DD, Fowler JH (2011): The evolution of overconfidence. *Nature*. 477:317-320.
15. Hume D (1900): *An enquiry concerning human understanding*. Chicago,: The Open court publishing co.; etc.
16. Helmholtz Hv (1867): *Handbuch der physiologischen Optik*. Leipzig,: Voss.
17. Feldman H, Friston KJ (2010): Attention, uncertainty, and free-energy. *Frontiers in human neuroscience*. 4:215.
18. Friston K, Kiebel S (2009): Predictive coding under the free-energy principle. *Philosophical transactions of the Royal Society of London Series B, Biological sciences*. 364:1211-1221.
19. Friston KJ, Stephan KE (2007): Free-energy and the brain. *Synthese*. 159:417-458.
20. Teufel C, Subramaniam N, Fletcher PC (2013): The role of priors in Bayesian models of perception. *Front Comput Neurosci*. 7:25.
21. McClelland JL, Rumelhart DE (1981): An interactive activation model of context effects in letter perception: I. An account of basic findings. *Psychological review*. .88:pp.
22. Rumelhart DE, McClelland JL (1982): An interactive activation model of context effects in letter perception: II. The contextual enhancement effect and some tests and extensions of the model. *Psychological review*. .89:pp.
23. Rao RP, Ballard DH (1999): Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature neuroscience*. 2:79-87.
24. Maia TV, Cleeremans A (2005): Consciousness: converging insights from connectionist modeling and neuroscience. *Trends in cognitive sciences*. 9:397-404.
25. Friston K (2005): A theory of cortical responses. *Philosophical transactions of the Royal Society of London Series B, Biological sciences*. 360:815-836.
26. Friston K (2005): Hallucinations and perceptual inferences. *The Behavioral and brain sciences*. 28:764-766.

27. Lupyan G, Swingle D (2012): Self-directed speech affects visual search performance. *Q J Exp Psychol (Hove)*. 65:1068-1085.
28. Lupyan G, Spivey MJ (2008): Perceptual processing is facilitated by ascribing meaning to novel stimuli. *Current biology : CB*. 18:R410-412.
29. Lupyan G (2008): The conceptual grouping effect: categories matter (and named categories matter more). *Cognition*. 108:566-577.
30. Lupyan G, Spivey MJ (2010): Making the invisible visible: verbal but not visual cues enhance visual detection. *PloS one*. 5:e11452.
31. Dikker S, Rabagliati H, Farmer TA, Pyllkanen L (2010): Early occipital sensitivity to syntactic category is based on form typicality. *Psychol Sci*. 21:629-634.
32. Dikker S, Rabagliati H, Pyllkanen L (2009): Sensitivity to syntax in visual cortex. *Cognition*. 110:293-321.
33. Jordan TR, Sheen M, AlJassmi MA, Paterson KB (2015): A New Demonstration of the Illusory Letters Phenomenon: Graphemic Restoration in Arabic Word Perception. *Perception*. 44:215-218.
34. Jordan TR, Thomas SM, Scott-Brown KC (1999): The illusory-letters phenomenon: an illustration of graphemic restoration in visual word recognition. *Perception*. 28:1413-1416.
35. Lupyan G, Thompson-Schill SL (2012): The evocative power of words: activation of concepts by verbal and nonverbal means. *J Exp Psychol Gen*. 141:170-186.
36. den Ouden HE, Friston KJ, Daw ND, McIntosh AR, Stephan KE (2009): A dual role for prediction error in associative learning. *Cerebral cortex*. 19:1175-1185.
37. Knill DC, Pouget A (2004): The Bayesian brain: the role of uncertainty in neural coding and computation. *Trends in neurosciences*. 27:712-719.
38. Swets J, Tanner WP, Jr., Birdsall TG (1961): Decision processes in perception. *Psychological review*. 68:301-340.
39. Scharf B, Quigley S, Aoki C, Peachey N, Reeves A (1987): Focused auditory attention and frequency selectivity. *Perception & psychophysics*. 42:215-223.
40. Marks LE, Wheeler ME (1998): Attention and the detectability of weak taste stimuli. *Chem Senses*. 23:19-29.
41. Esquirol J (1814): Hallucinations. *Dictionnaire des sciences médicales*. Paris: CLF Panckouche. 1822.
42. Bleuler E (2013): *Lehrbuch der psychiatrie*. Springer-Verlag.
43. Winokur G, Scharfetter C, Angst J (1985): The diagnostic value in assessing mood congruence in delusions and hallucinations and their relationship to the affective state. *Eur Arch Psychiatry Neurol Sci*. 234:299-302.
44. Schneider F, Gur RC, Gur RE, Shtasel DL (1995): Emotional processing in schizophrenia: neurobehavioral probes in relation to psychopathology. *Schizophrenia research*. 17:67-75.
45. Olson PR, Suddeth JA, Peterson PJ, Egelhoff C (1985): Hallucinations of widowhood. *J Am Geriatr Soc*. 33:543-547.
46. Dewi Rees W (1971): The hallucinations of widowhood. *Br Med J*. 4:37-41.
47. Baethge C (2002): Grief hallucinations: true or pseudo? Serious or not? An inquiry into psychopathological and clinical features of a common phenomenon. *Psychopathology*. 35:296-302.
48. van Os J, Linscott RJ, Myin-Germeys I, Delespaul P, Krabbendam L (2009): A systematic review and meta-analysis of the psychosis continuum: evidence for a psychosis proneness-persistence-impairment model of psychotic disorder. *Psychological medicine*. 39:179-195.
49. Galdos M, Simons C, Fernandez-Rivas A, Wichers M, Peralta C, Lataster T, et al. (2011): Affectively salient meaning in random noise: a task sensitive to psychosis liability. *Schizophrenia bulletin*. 37:1179-1186.
50. Diederer KM, van Lutterveld R, Sommer IE (2012): Neuroimaging of voice hearing in non-psychotic individuals: a mini review. *Frontiers in human neuroscience*. 6:111.

51. Laroi F, Sommer IE, Blom JD, Fernyhough C, Ffytche DH, Hugdahl K, et al. (2012): The characteristic features of auditory verbal hallucinations in clinical and nonclinical groups: state-of-the-art overview and future directions. *Schizophrenia bulletin*. 38:724-733.
52. Johns LC, Kompus K, Connell M, Humpston C, Lincoln TM, Longden E, et al. (2014): Auditory verbal hallucinations in persons with and without a need for care. *Schizophrenia bulletin*. 40 Suppl 4:S255-264.
53. de Leede-Smith S, Barkus E (2013): A comprehensive review of auditory verbal hallucinations: lifetime prevalence, correlates and mechanisms in healthy and clinical individuals. *Front Hum Neurosci*. 7:367.
54. Johns LC, Nazroo JY, Bebbington P, Kuipers E (2002): Occurrence of hallucinatory experiences in a community sample and ethnic variations. *Br J Psychiatry*. 180:174-178.
55. Teufel C, Subramaniam N, Dobler V, Perez J, Finnemann J, Mehta PR, et al. (2015): Shift toward prior knowledge confers a perceptual advantage in early psychosis and psychosis-prone healthy individuals. *Proc Natl Acad Sci U S A*.
56. Ellson DG (1941): Hallucinations produced by sensory conditioning. *Journal of Experimental Psychology*: American Psychological Association, pp 1.
57. Kot T, Serper M (2002): Increased susceptibility to auditory conditioning in hallucinating schizophrenic patients: a preliminary investigation. *J Nerv Ment Dis*. 190:282-288.
58. Warburton DM, Wesnes K, Edwards J, Larrad D (1985): Scopolamine and the sensory conditioning of hallucinations. *Neuropsychobiology*. 14:198-202.
59. Agathon M, Roussel A (1973): [Use of a sensory conditioning test in patients treated with psychotropic drugs]. *Int Pharmacopsychiatry*. 8:221-233.
60. Brogden WJ (1947): Sensory pre-conditioning of human subjects. *J Exp Psychol*. 37:527-539.
61. Whitson JA, Galinsky AD (2008): Lacking control increases illusory pattern perception. *Science*. 322:115-117.
62. Sutton R (1992): *Proceedings of the 7th Yale Workshop on Adaptive and Learning Systems*. New Haven, Connecticut: Yale University Press, pp 161-166.
63. Anderson BDO, Moore, J. B. (1979): *Optimal Filtering*. Englewood Cliffs, New Jersey: Prentice-Hall.
64. Dayan P, Kakade S, Montague PR (2000): Learning and selective attention. *Nature neuroscience*. 3 Suppl:1218-1223.
65. Iglesias S, Mathys C, Brodersen KH, Kasper L, Piccirelli M, den Ouden HE, et al. (2013): Hierarchical prediction errors in midbrain and basal forebrain during sensory learning. *Neuron*. 80:519-530.
66. Park HJ, Friston K (2013): Structural and functional brain networks: from connections to cognition. *Science*. 342:1238411.
67. Allen M, Fardo F, Dietz MJ, Hillebrandt H, Friston KJ, Rees G, et al. (2015): Anterior insula coordinates hierarchical processing of tactile mismatch responses. *NeuroImage*. 127:34-43.
68. Shergill SS, Brammer MJ, Amaro E, Williams SC, Murray RM, McGuire PK (2004): Temporal course of auditory hallucinations. *The British journal of psychiatry : the journal of mental science*. 185:516-517.
69. Sommer IE, Diederer KM, Blom JD, Willems A, Kushan L, Slotema K, et al. (2008): Auditory verbal hallucinations predominantly activate the right inferior frontal area. *Brain : a journal of neurology*. 131:3169-3177.
70. Geddes MR, Tie Y, Gabrieli JD, McGinnis SM, Golby AJ, Whitfield-Gabrieli S (2016): Altered functional connectivity in lesional peduncular hallucinosis with REM sleep behavior disorder. *Cortex; a journal devoted to the study of the nervous system and behavior*. 74:96-106.
71. Hurley LM, Devilbiss DM, Waterhouse BD (2004): A matter of focus: monoaminergic modulation of stimulus coding in mammalian sensory networks. *Curr Opin Neurobiol*. 14:488-495.
72. Yu AJ, Dayan P (2005): Uncertainty, neuromodulation, and attention. *Neuron*. 46:681-692.
73. Servan-Schreiber D, Printz H, Cohen JD (1990): A network model of catecholamine effects: gain, signal-to-noise ratio, and behavior. *Science*. 249:892-895.
74. Eldar E, Cohen JD, Niv Y (2013): The effects of neural gain on attention and learning. *Nature neuroscience*. 16:1146-1153.

75. Summerfield C, Koechlin E (2008): A neural representation of prior information during perceptual inference. *Neuron*. 59:336-347.
76. Blanke O, Landis T, Seeck M (2000): Electrical cortical stimulation of the human prefrontal cortex evokes complex visual hallucinations. *Epilepsy & behavior : E&B*. 1:356-361.
77. Mc LT, Meyer A (1949): Anatomical correlates of improvement after leucotomy. *The Journal of mental science*. 95:182-196.
78. Gratton C, Nomura EM, Perez F, D'Esposito M (2012): Focal brain lesions to critical locations cause widespread disruption of the modular organization of the brain. *Journal of cognitive neuroscience*. 24:1275-1285.
79. van den Heuvel MP, Sporns O (2011): Rich-club organization of the human connectome. *The Journal of neuroscience : the official journal of the Society for Neuroscience*. 31:15775-15786.
80. Whitfield-Gabrieli S, Thermenos HW, Milanovic S, Tsuang MT, Faraone SV, McCarley RW, et al. (2009): Hyperactivity and hyperconnectivity of the default network in schizophrenia and in first-degree relatives of persons with schizophrenia. *Proc Natl Acad Sci U S A*. 106:1279-1284.
81. Carhart-Harris RL, Friston KJ (2010): The default-mode, ego-functions and free-energy: a neurobiological account of Freudian ideas. *Brain : a journal of neurology*. 133:1265-1283.
82. van den Heuvel MP, Sporns O, Collin G, Scheewe T, Mandl RC, Cahn W, et al. (2013): Abnormal rich club organization and functional brain dynamics in schizophrenia. *JAMA psychiatry*. 70:783-792.
83. Chanes L, Barrett LF (2016): Redefining the Role of Limbic Areas in Cortical Processing. *Trends in cognitive sciences*. 20:96-106.
84. Dehaene S, Kerszberg M, Changeux JP (1998): A neuronal model of a global workspace in effortful cognitive tasks. *Proceedings of the National Academy of Sciences of the United States of America*. 95:14529-14534.
85. Dehaene S, Changeux JP (2011): Experimental and theoretical approaches to conscious processing. *Neuron*. 70:200-227.
86. Fodor J (1983): *The Modularity of Mind: An essay on faculty psychology*. Harvard: MIT Press.
87. Friston KJ, Price CJ (2011): Modules and brain mapping. *Cognitive neuropsychology*. 28:241-250.
88. Colombo M, Colombo M (2013): Moving Forward (and beyond) the Modularity Debate: A Network Perspective. *Philosophy of Science*. 80 (3).
89. Penny WD, Stephan KE, Mechelli A, Friston KJ (2004): Comparing dynamic causal models. *NeuroImage*. 22:1157-1172.
90. Curcic-Blake B, Liemburg E, Vercammen A, Swart M, Knegtering H, Bruggeman R, et al. (2013): When Broca goes uninformed: reduced information flow to Broca's area in schizophrenia patients with auditory hallucinations. *Schizophrenia bulletin*. 39:1087-1095.
91. Barrett LF, Simmons WK (2015): Interoceptive predictions in the brain. *Nature reviews Neuroscience*. 16:419-429.
92. van den Heuvel MP, Sporns O (2013): An anatomical substrate for integration among functional networks in human cortex. *The Journal of neuroscience : the official journal of the Society for Neuroscience*. 33:14489-14500.
93. Hohwy J, Roepstorff A, Friston K (2008): Predictive coding explains binocular rivalry: an epistemological review. *Cognition*. 108:687-701.
94. Einhauser W, Stout J, Koch C, Carter O (2008): Pupil dilation reflects perceptual selection and predicts subsequent stability in perceptual rivalry. *Proceedings of the National Academy of Sciences of the United States of America*. 105:1704-1709.
95. Hoffman RE, Quinlan DM, Mazure CM, McGlashan TM (2001): Cortical instability and the mechanism of mania: a neural network simulation and perceptual test. *Biological psychiatry*. 49:500-509.
96. McBain R, Norton DJ, Kim J, Chen Y (2011): Reduced cognitive control of a visually bistable image in schizophrenia. *Journal of the International Neuropsychological Society : JINS*. 17:551-556.
97. Tononi G, Edelman GM (2000): Schizophrenia and the mechanisms of conscious integration. *Brain research Brain research reviews*. 31:391-400.

98. Tononi G, Sporns O, Edelman GM (1994): A measure for brain complexity: relating functional segregation and integration in the nervous system. *Proceedings of the National Academy of Sciences of the United States of America*. 91:5033-5037.
99. Lupyan G, Clark A (2015): Words and the World: Predictive Coding and the Language-Perception-Cognition Interface. *Current Directions in Psychological Science*. 24:279-284.
100. Dakin S, Carlin P, Hemsley D (2005): Weak suppression of visual context in chronic schizophrenia. *Current biology : CB*. 15:R822-824.
101. Koethe D, Gerth CW, Neatby MA, Haensel A, Thies M, Schneider U, et al. (2006): Disturbances of visual information processing in early states of psychosis and experimental delta-9-tetrahydrocannabinol altered states of consciousness. *Schizophrenia research*. 88:142-150.
102. Teufel C, Subramaniam N, Dobler V, Perez J, Finnemann J, Mehta PR, et al. (2015): Shift toward prior knowledge confers a perceptual advantage in early psychosis and psychosis-prone healthy individuals. *Proceedings of the National Academy of Sciences of the United States of America*. 112:13401-13406.
103. Norton D, Ongur D, Stromeyer C, 3rd, Chen Y (2008): Altered 'three-flash' illusion in response to two light pulses in schizophrenia. *Schizophrenia research*. 103:275-282.
104. Ford JM (2016): Studying auditory verbal hallucinations using the RDoC framework. *Psychophysiology*. 53:298-304.
105. Adams RA, Stephan KE, Brown HR, Frith CD, Friston KJ (2013): The computational anatomy of psychosis. *Frontiers in psychiatry*. 4:47.
106. Powers III AR, Gancsos MG, Finn ES, Morgan PT, Corlett PR (2015): Ketamine-Induced Hallucinations. *Psychopathology*. 48:376-385.
107. Corlett PR, Frith CD, Fletcher PC (2009): From drugs to deprivation: a Bayesian framework for understanding models of psychosis. *Psychopharmacology*. 206:515-530.
108. Horga G, Schatz KC, Abi-Dargham A, Peterson BS (2014): Deficits in predictive coding underlie hallucinations in schizophrenia. *The Journal of neuroscience : the official journal of the Society for Neuroscience*. 34:8072-8082.
109. Notredame CE, Pins D, Deneve S, Jardri R (2014): What visual illusions teach us about schizophrenia. *Frontiers in integrative neuroscience*. 8:63.
110. Arieti S (1974): An overview of schizophrenia from a predominantly psychological approach. *The American journal of psychiatry*. 131:241-249.
111. Hoffman RE (2010): Revisiting Arieti's "listening attitude" and hallucinated voices. *Schizophrenia bulletin*. 36:440-442.
112. Jardri R, Deneve S (2013): Circular inferences in schizophrenia. *Brain : a journal of neurology*. 136:3227-3241.
113. Collerton D, Perry E, McKeith I (2005): Why people see things that are not there: a novel Perception and Attention Deficit model for recurrent complex visual hallucinations. *The Behavioral and brain sciences*. 28:737-757; discussion 757-794.
114. Phillips WA, Clark A, Silverstein SM (2015): On the functions, mechanisms, and malfunctions of intracortical contextual modulation. *Neuroscience and biobehavioral reviews*. 52:1-20.
115. Corlett PR, Fletcher, P.C. (2014): Computational Psychiatry: A Rosetta Stone linking the brain to mental illness. *Lancet Psychiatry*.
116. Sass LA (1994): *Paradoxes of delusion: Wittgenstein, Schreber, and the schizophrenic mind*. Ithaca: Cornell University Press.
117. Mikles NL, Laycock JP (2015): Tracking the Tulpa. *Nova Religio: The Journal of Alternative and Emergent Religions*. 19:87-97.
118. Cichy RM, Heinze J, Haynes JD (2012): Imagery and perception share cortical representations of content and location. *Cerebral cortex*. 22:372-380.
119. Ffytche DH (2005): Visual hallucinations and the Charles Bonnet syndrome. *Current psychiatry reports*. 7:168-179.

120. Seth AK, Suzuki K, Critchley HD (2011): An interoceptive predictive coding model of conscious presence. *Frontiers in psychology*. 2:395.
121. Manita S, Suzuki T, Homma C, Matsumoto T, Odagawa M, Yamada K, et al. (2015): A Top-Down Cortical Circuit for Accurate Sensory Perception. *Neuron*. 86:1304-1316.
122. (!!! INVALID CITATION !!! (107)).
123. van Leeuwen TM, Singer W, Nikolic D (2015): The Merit of Synesthesia for Consciousness Research. *Frontiers in psychology*. 6:1850.

Figure 1. Hubs, Connectors and Modules.

Connectors (black circles) and hubs (white squares) in modules. Hubs have many within-module connections. Connectors have many between module connections

