HHS Public Access

Author manuscript

Science. Author manuscript; available in PMC 2018 February 11.

Published in final edited form as:

Science. 2017 August 11; 357(6351): 596–600. doi:10.1126/science.aan3458.

Pavlovian Conditioning-Induced Hallucinations Result from Overweighting of Perceptual Priors

A.R. Powers[1], **C. Mathys**[2,3,4], and **P.R. Corlett**[1,*]

[1]Department of Psychiatry, Yale University School of Medicine, New Haven, CT, USA [2]International School for Advanced Studies (SISSA), Trieste, Italy [3]Max Planck UCL Centre for Computational Psychiatry and Ageing Research, London, UK [4]Translational Neuromodeling Unit (TNU), Institute for Biomedical Engineering, University of Zurich and ETH Zurich, Zurich, Switzerland

Abstract

Some people hear voices that others do not, but only some of those people seek treatment. Using a Pavlovian learning task, we induced conditioned hallucinations in four groups of people who differed orthogonally in their voice-hearing and treatment-seeking statuses. People who hear voices were significantly more susceptible to the effect. Using functional neuroimaging and computational modeling of perception, we identified processes that differentiated voice-hearers from non-voice-hearers and treatment-seekers from non-treatment-seekers and characterized a brain circuit that mediated the conditioned hallucinations. These data demonstrate the profound and sometimes pathological impact of top-down cognitive processes on perception and may represent an objective means to discern people with a need for treatment from those without.

Perception is not simply the passive reception of inputs (1). We actively infer the causes of our sensations (2). These inferences are influenced by our prior experiences (3). Priors and inputs might be combined according to Bayes' rule (4). Prediction errors, the mismatch between priors and inputs, contribute to belief updating (5). Hallucinations (percepts without external stimulus) may arise when strong priors cause a percept in the absence of input (6). We tested this theory by engendering new priors about auditory stimuli in human observers using Pavlovian conditioning.

Even in healthy individuals, the repeated co-occurrence of visual and auditory stimuli can induce auditory hallucinations (7). We examined this effect with functional imaging. Some argue that, in patients with psychosis, weak priors lead to aberrant prediction errors,

Corresponding author. philip.corlett@yale.edu.

Conflicts of Interest: The authors declare no conflicts of interest. Model code and data stored at ModelDB (http://senselab.med.yale.edu/ModelDB/showModel.cshtml?model=229278) Imaging data stored at NeuroVault (/collections/OCFEJCQE/).

List of Supplementary Materials:
1) Materials and Methods
2) Supplemental Results Text
3) Tables S1–S8
4) Figures S1–S7
5) References

resulting in auditory verbal hallucinations (AVH) (8). Others have observed strong priors in patients, but the effects were not specific to hallucinations (9, 10). Such inconsistencies may reflect the hierarchical organization of perception; perturbations may impact some levels of the hierarchy and not others (9). We used computational modeling to infer the strength of participants' hierarchical perceptual beliefs from their behavioral responses during conditioning (11). Importantly, our model captured how priors are combined with sensory evidence, allowing us to directly test the strong prior hypothesis.

Participants worked to detect a 1-kHz tone occurring concurrently with presentation of a checkerboard visual stimulus. First, we determined individual thresholds for detection and psychometric curves (12). Then, at the start of conditioning, the tone was presented frequently at threshold (Fig. 1A, left), engendering a belief in audio-visual association. This belief was then tested (Fig. 1A, right) with increasingly frequent sub-threshold and target-absent trials (Fig. 1B). *Conditioned hallucinations* occurred when subjects reported tones that were not presented, conditional upon the visual stimulus.

We recruited four groups of subjects (Fig. 1C): people with a diagnosed psychotic illness who heard voices (P+H+, n=15); those with similar who did not hear voices (P+H−, n=14); an active control group who heard daily voices, but had no diagnosed illness (13) (P-H+, n=15; they attributed their experiences metaphyscially (14), see Supplement); and finally, controls without diagnosis or voices (P-H−, n=15).

Groups were matched demographically (Tables S1–S4). Rates of detection of tones at threshold were similar across groups. All groups demonstrated conditioned hallucinations. However, those with daily hallucinations endorsed more conditioned hallucinations than those without, regardless of diagnosis (Fig. 1D; $F_{1,55}=19.59$; $p=5.82\times10^{-5}$). This effect remained after accounting for differences in detection thresholds (Fig.1E, Fig. S1, Table S5). Group differences in propensity to report tones were observed only in the No-Tone and 25% Likelihood of Detection conditions (Fig. 1F; intensity-by-hallucination status $F_{3,165}=13.59$, $p=5.73\times10^{-4}$).

Participants also rated their decision confidence by holding down the response button (Fig. 1G). Participant confidence varied with stimulus intensity ("yes": $R=0.39$; $p=7.46\times10^{-10}$; "no": $R=0.22$; $p=9.02\times10^{-4}$). However, hallucinators were more confident in their conditioned hallucinations than non-hallucinators ($F_{1,53}=6.50$; $p=0.045$). Both conditioned hallucinations and confidence correlated with hallucination severity outside of the laboratory (Fig. 1H–I; Fig. S3).

In order to establish whether conditioned hallucinations involved true percepts, we first identified tone-responsive regions from thresholding runs (Fig. 2A; peaks at [−60 −20 2] and [62 −28 10]). As observed with elementary hallucinations (15), activity in tone-responsive regions was greater during conditioned hallucinations compared to correct rejections (Fig. 2B; $t_{56}=4.93$, $p=7.59\times10^{-6}$). Electrical stimulation of this region in human patients produces AVH (16). Taken together, these findings are consistent conditioned hallucinations involving actual perception.

Whole-brain analysis revealed that conditioned hallucinations also engaged anterior insula cortex (AIC), inferior frontal gyrus, head of caudate, anterior cingulate cortex (ACC), auditory cortex, and posterior superior temporal sulcus (STS) (Fig. 2C, Table S6). A meta-analysis of symptom-capture-based studies examining neural activity of AVH highlighted similar regions (17) (Fig. 2D). AIC and ACC responses frequently correlate with stimulus salience (18). However, their activation prior to near-threshold stimulus presentation predicts detection (19). Caudate is engaged during audiovisual associative learning (20). Likewise, AIC and ACC are engaged during multisensory integration (21).

There were no significant between-group differences in brain responses during conditioned hallucinations. However, hallucinators deactivated ACC more (peak at [−16, 54, 14]; cluster-extent thresholded, starting value 0.005, critical $k_e = 99$) during correct rejections compared to non-hallucinators (Fig. 2E–F).

To further dissect conditioned hallucinations we modeled their underlying computational mechanisms (Fig 3A) using the Hierarchical Gaussian Filter (HGF (11)). We defined a perceptual model consisting of low-level perceptual beliefs (X_1), visual-auditory associations (X_2), and the volatility of those associations (X_3), as well as learning rates encoding the relationships between levels (,). Critically, our perceptual model allowed for variability in weighting between sensory evidence and perceptual beliefs (ν). For, prior and observation have equal weight; for the prior has more weight than the observation (strong priors); and for the observation has more weight than the prior (weak priors). The resultant posterior probability of a tone is then fed to a separate response model.

Model parameters were fit to behavioral data and the model was optimized using log model evidence and simulations of observed behavior (Figs. S3 and S4). Mean trajectories of perceptual beliefs were compared across groups (Fig. 3B–D). Participants with hallucinations exhibited stronger beliefs at layers 1 (Fig. 3D) and 2 (Fig. 3C; X_1:$F_{11,605}=4.8$,$p=3.89\times10^{-7}$; X_2: $F_{11,605}=3.89$,$p=1.84\times10^{-5}$). X_3 beliefs evolved less in those with psychosis, who failed to recognize the increasing volatility in contingencies (Fig. 3A; $F_{11,605}=2.11$,$p=0.018$).

Consistent with strong-prior theory, ν was significantly larger in those with hallucinations when compared to their non-hallucinating counterparts (Fig. 3E), regardless of diagnosis ($F_{1,55}=13.96$,$p=4.45\times10^{-4}$). Response model parameters did not differ across the groups (Fig. 3F).

We regressed model parameters onto task-induced brain responses (Fig. 4A). The X_1 trajectory co-varied with several conditioned hallucination-responsive regions including STS (Table S7). X_3 trajectories, by contrast, covaried with hippocampus/parahippocampal gyrus and medial cerebellum (Table S8). Parameter estimates from the X_1 sensitive STS (Fig. 4B; [−46−36,0],$T_{57}=2.09$,$p=0.042$) and AIC (Fig. 4C; [36,8,−8],$T_{57}=2.26$,$p=0.027$) were significantly greater in those with hallucinations versus those without. This is consistent with STS conferring auditory expectations that are responsive to incoming visual input (22). Parameter estimates from the X_3 responsive cerebellar vermis (Fig. 4D; [−2,−52,−16]) were lower in participants with psychosis compared to those without ($T_{57}=2.05$,$p=0.045$). In the

model, subjects with psychosis were significantly less sensitive to the changes in contingency as the task progressed. Psychotic symptoms are often associated with pathological rigidity. Belief updating correlated with responses in the hippocampus and cerebellum. Hippocampal activity correlates with uncertainty in perceptual predictions (23). The cerebellum has likewise been associated with production and updating of predictive models (24).

Our X_1, X_2, and ν findings are consistent with a strong prior theory of hallucinations. The X_3 findings in psychotic patients may reflect a strong prior that contingencies are fixed. On the other hand, they could reflect a weak prior on volatility. These beliefs were not associated with hallucinations but rather psychosis more broadly. Under chronic uncertainty, secondary to consistent belief violation, it may be adaptive to resist updating beliefs (25).

Consistent with previous work applying signal detection theory (SDT) to AVH (26), we found liberal criteria and low perceptual sensitivity in our H+ groups. A liberal criterion may reflect poor reality monitoring (26).

However, meta-d' (a metric of participants' meta-cognitive sensitivity) did not differ significantly between groups (Fig. S6). SDT is a descriptive tool that does not distinguish aberrant perceptions from decisions. Our modeling work, however, localized group differences to the perceptual model alone. The prior weighting parameter (ν) distinguished H+ from H− groups and also predicted confidence in conditioned hallucinations (Fig. S7). Our observations support a strong perceptual prior explanation of hallucinations. They suggest precision treatments for hallucinations, like targeting cholinergically mediated priors (27) and interventions to mollify psychosis more broadly, like cerebellar transcranial magnetic stimulation (28).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors dedicate this work to the memory and legacy of Ralph E. Hoffman, M.D. Additional thanks to Megan Kelley, Adina Bianchi, Shivani Bhatt, and Erin Feeney for technical assistance as well as Drs. Larry Marks, Scott Woods, and John Krystal for their advice. 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, NIMH Grant 5R01MH067073-09, 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), NIH roadmap for Medical Research, the Clinical Neurosciences Division, U.S. Department of Veterans Affairs, National Center for Post-Traumatic Stress Disorders, VACHS, West Haven, CT, USA. 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. ARP was supported by the Integrated Mentored Patient-Oriented Research Training (IMPORT) in Psychiatry grant (5R25MH071584-07) as well as the Clinical Neuroscience Research Training in Psychiatry grant (5T32MH19961-14) from the NIMH and a VA Schizophrenia Research Special Fellowship, VACHS, West Haven, CT, USA. Additional support was provided by the Yale Detre Fellowship for Translational Neuroscience as well as the Brain and Behavior Research Foundation in the form of a NARSAD Young Investigator Award for Dr. Powers.

References and Notes

1. Helmholz, Hv. Treatise on physiological optics. Voss, Hamburg, 3rd, editor. 1909.

2. Friston K. A theory of cortical responses. Philosophical transactions of the Royal Society of London Series B, Biological sciences. 2005; 360:815–836. [PubMed: 15937014]
3. Rao RP, Ballard DH. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. Nature neuroscience. 1999; 2:79–87. [PubMed: 10195184]
4. Bayes T. An Essay Towards Solving a Problem in the Doctrine of Chances. Biometrika. 1958; 45:296–315.
5. Adams RA, Stephan KE, Brown HR, Frith CD, Friston KJ. The computational anatomy of psychosis. Frontiers in psychiatry. 2013; 4:47. [PubMed: 23750138]
6. Friston KJ. Hallucinations and perceptual inference. Behavioral and Brain Sciences. 2005; 28:764.
7. Ellison DG. Hallucinations produced by sensory conditioning. J Exp Psychol. 1941; 28:1–20.
8. Horga G, Schatz KC, Abi-Dargham A, Peterson BS. Deficits in predictive coding underlie hallucinations in schizophrenia. The Journal of neuroscience : the official journal of the Society for Neuroscience. 2014; 34:8072–8082. [PubMed: 24920613]
9. Teufel C, et al. 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. 2015; 112:13401–13406. [PubMed: 26460044]
10. Schmack K, Rothkirch M, Priller J, Sterzer P. Enhanced predictive signalling in schizophrenia. Human brain mapping. 2017
11. Mathys C, Daunizeau J, Friston KJ, Stephan KE. A bayesian foundation for individual learning under uncertainty. Frontiers in human neuroscience. 2011; 5:39. [PubMed: 21629826]
12. Watson AB, Pelli DG. QUEST: a Bayesian adaptive psychometric method. Percept Psychophys. 1983; 33:113–120. [PubMed: 6844102]
13. Verdoux H, van Os J. Psychotic symptoms in non-clinical populations and the continuum of psychosis. Schizophrenia research. 2002; 54:59–65. [PubMed: 11853979]
14. Powers AR 3rd, Kelley MS, Corlett PR. Varieties of Voice-Hearing: Psychics and the Psychosis Continuum. Schizophr Bull. 2017; 43:84–98. [PubMed: 28053132]
15. Pearson J, et al. Sensory dynamics of visual hallucinations in the normal population. eLife. 2016; 5
16. Penfield W, Perot P. The Brain's Record of Auditory and Visual Experience. A Final Summary and Discussion. Brain : a journal of neurology. 1963; 86:595–696. [PubMed: 14090522]
17. Zmigrod L, Garrison JR, Carr J, Simons JS. The neural mechanisms of hallucinations: A quantitative meta-analysis of neuroimaging studies. Neuroscience and biobehavioral reviews. 2016; 69:113–123. [PubMed: 27473935]
18. Sterzer P, Kleinschmidt A. Anterior insula activations in perceptual paradigms: often observed but barely understood. Brain Struct Funct. 2010; 214:611–622. [PubMed: 20512379]
19. Sadaghiani S, Hesselmann G, Kleinschmidt A. Distributed and antagonistic contributions of ongoing activity fluctuations to auditory stimulus detection. The Journal of neuroscience : the official journal of the Society for Neuroscience. 2009; 29:13410–13417. [PubMed: 19846728]
20. den Ouden HE, Friston KJ, Daw ND, McIntosh AR, Stephan KE. A dual role for prediction error in associative learning. Cerebral cortex. 2009; 19:1175–1185. [PubMed: 18820290]
21. Laurienti PJ, et al. Cross-modal sensory processing in the anterior cingulate and medial prefrontal cortices. Hum Brain Mapp. 2003; 19:213–223. [PubMed: 12874776]
22. Powers AR 3rd, Hevey MA, Wallace MT. Neural correlates of multisensory perceptual learning. J Neurosci. 2012; 32:6263–6274. [PubMed: 22553032]
23. Schiffer AM, Ahlheim C, Wurm MF, Schubotz RI. Surprised at all the entropy: hippocampal, caudate and midbrain contributions to learning from prediction errors. PloS one. 2012; 7:e36445. [PubMed: 22570715]
24. Shergill SS, et al. Functional magnetic resonance imaging of impaired sensory prediction in schizophrenia. JAMA psychiatry. 2014; 71:28–35. [PubMed: 24196370]
25. Karlsson MP, Tervo DG, Karpova AY. Network resets in medial prefrontal cortex mark the onset of behavioral uncertainty. Science. 2012; 338:135–139. [PubMed: 23042898]
26. Bentall RP, Slade PD. Reality testing and auditory hallucinations: a signal detection analysis. The British journal of clinical psychology / the British Psychological Society. 1985; 24(Pt 3):159–169.

27. Warburton DM, Wesnes K, Edwards J, Larrad D. Scopolamine and the sensory conditioning of hallucinations. Neuropsychobiology. 1985; 14:198–202. [PubMed: 3835496]

28. Parker KL, Narayanan NS, Andreasen NC. The therapeutic potential of the cerebellum in schizophrenia. Frontiers in systems neuroscience. 2014; 8:163. [PubMed: 25309350]

29. Treutwein B, Strasburger H. Fitting the psychometric function. Percept Psychophys. 1999; 61:87–106. [PubMed: 10070202]

30. Ravicz ME, Melcher JR, Kiang NY. Acoustic noise during functional magnetic resonance imaging. J Acoust Soc Am. 2000; 108:1683–1696. [PubMed: 11051496]

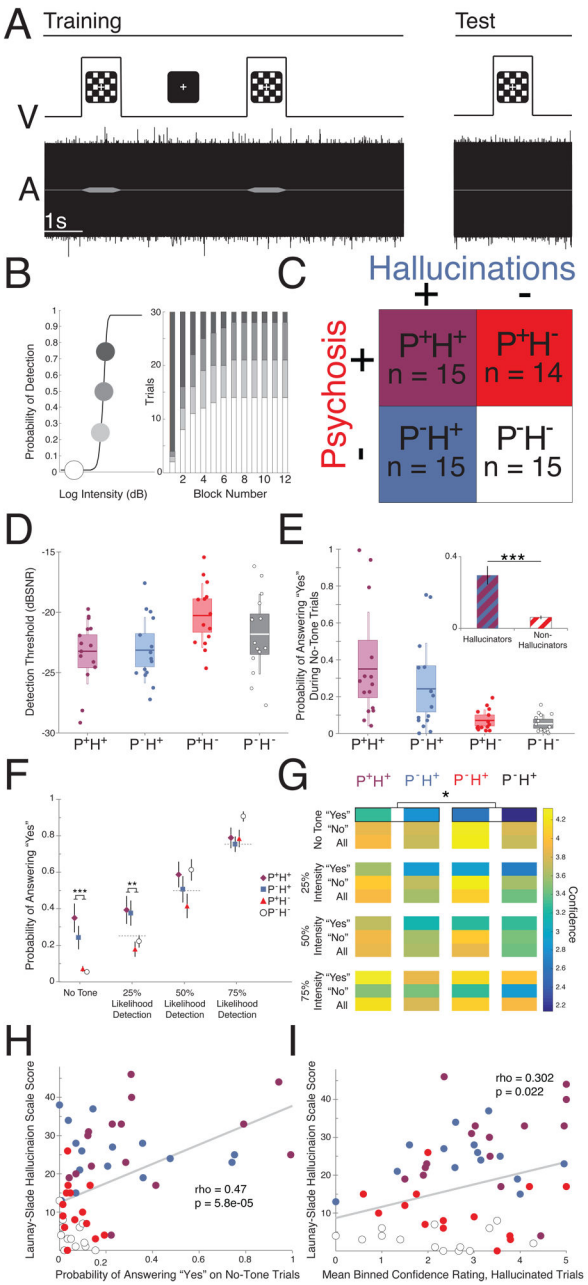

Fig 1. Methods and Behavioral Results

A. Trials consisted of simultaneous presentation of a 1000-Hz tone in white noise and a visual checkerboard. B. We estimated individual psychometric curves for tone detection (left) and then systematically varied stimulus intensity over twelve blocks of 30 conditioning trials. Threshold tones were more likely early and absent tones were more likely later (right). C. Groups varied along two dimensions: the presence (+) or absence (-) of daily AVH (blue) and the presence (+) or absence (−) of a diagnosable psychotic-spectrum illness (red). D. Detection thresholds. Error bars represent ±1 SD, boxes represent ± 1 SEM. E. Probability of conditioned hallucinations varied according to hallucination status. Main panel: error bars represent ±1 SD, boxes represent ±1 SEM. Inset: error bars represent ±1 SEM. F.

Differences between hallucinating and non-hallucinating groups were found only in the target-absent and 25% Likelihood of Detection conditions. Error bars represent ±1 SEM. G. Hallucinators were more confident than non-hallucinators when reporting a tone that did not exist. H–I. Both the probability of reporting conditioned hallucinations (H) and the confidence with which they were reported (I) correlated with a measure of hallucination severity.

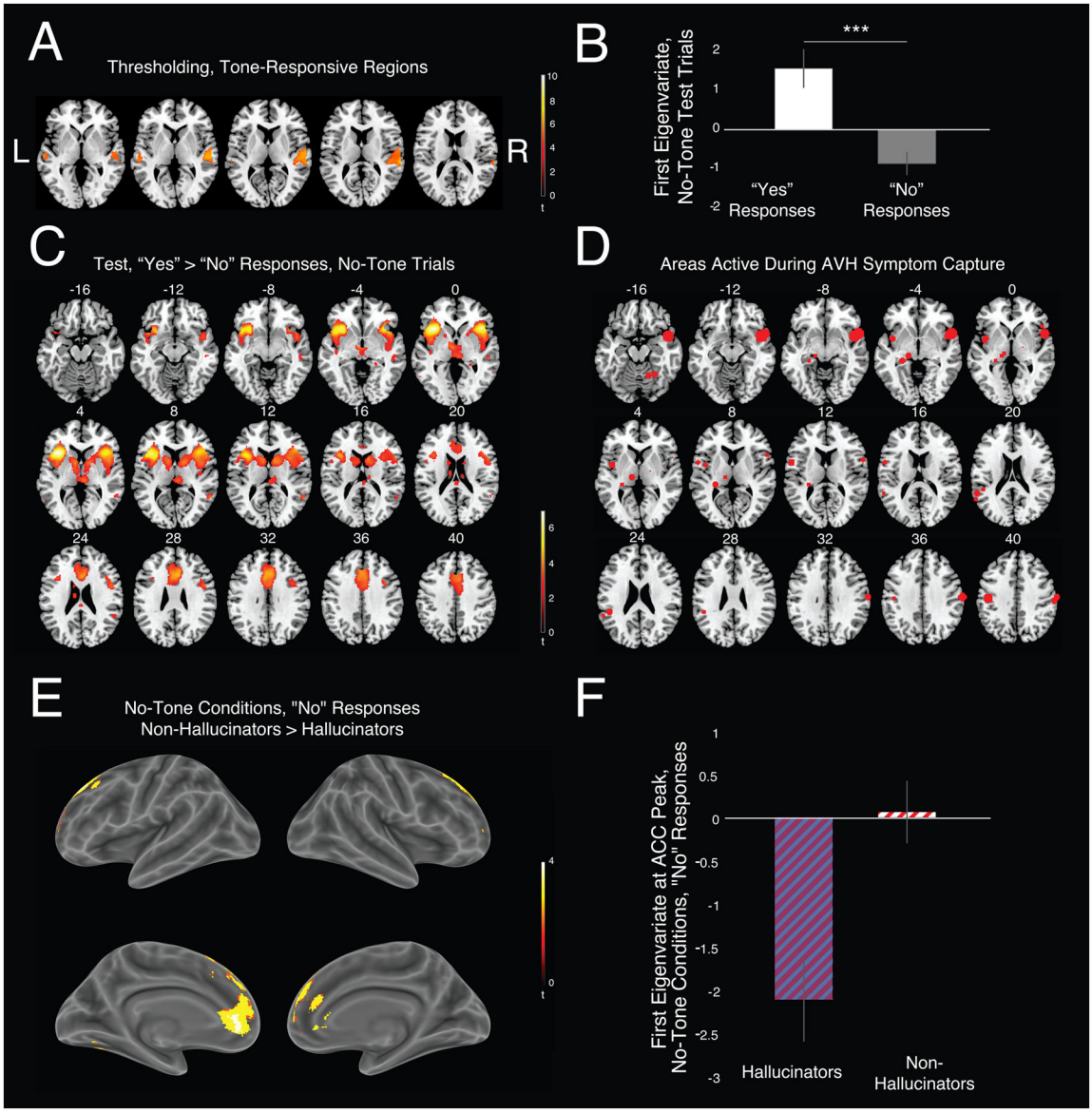

Fig 2. Imaging Results

A. Bilateral supplemental auditory cortex co-varied with tone intensity during thresholding (FWE-corrected, $P < 0.05$). B. Parameter estimates from this region showed increased activation during conditioned hallucinations. C. Whole-brain analysis during conditioned hallucinations (FDR-corrected, $P < 0.05$). D. Clusters derived from a meta-analysis (17) of AVH experiences during functional imaging. E–F. Hallucinators were much less likely to engage anterior cingulate cortex during correct rejections. Error bars represent ±1 SEM.

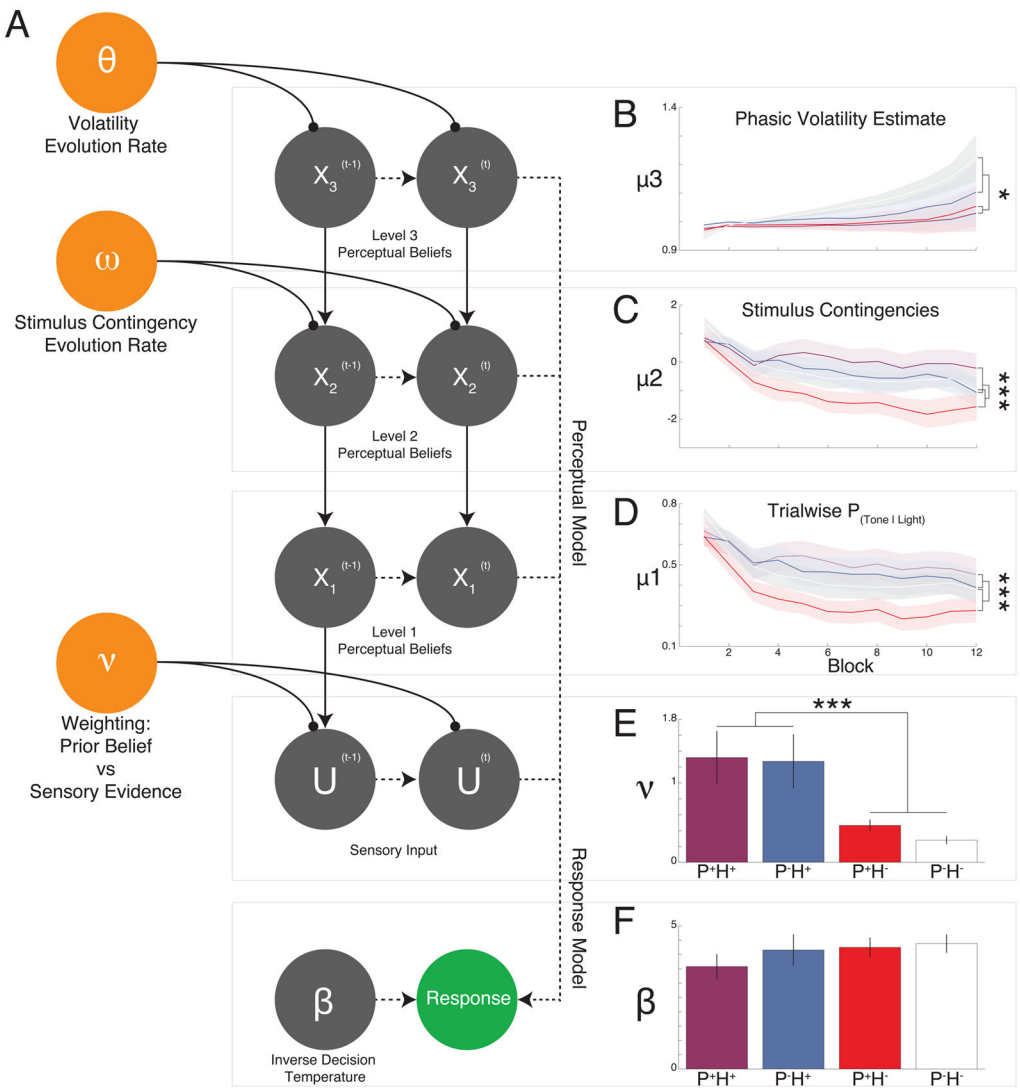

Fig 3. Hierarchical Gaussian Filter Analysis

A. Computational model, mapping from experimental stimuli to observed responses through perceptual and response models. The first level (X_1) represents whether the subject believes a tone was present or not on trial t. The second level (X_2) is their belief that visual cues are associated with tones. The third level (X_3) is their belief about the volatility of the second level. The HGF allows for individual variability in weighting between sensory evidence and perceptual beliefs (parameter ν). B. At X_3 there was a significant block-by-psychosis interaction. C–D. Significant block-by-hallucination status interactions were seen at layers X_1 (D) and X_2 (C). E. Nu (ν), was significantly higher in those with hallucinations when compared to their non-hallucinating counterparts. F. No main effects of group or interaction effects were seen for the decision noise parameter within the response model. Error bars and line shadings represent ±1 SEM. P+H+: purple; P-H+: blue; P+H−: red; P-H−: white.

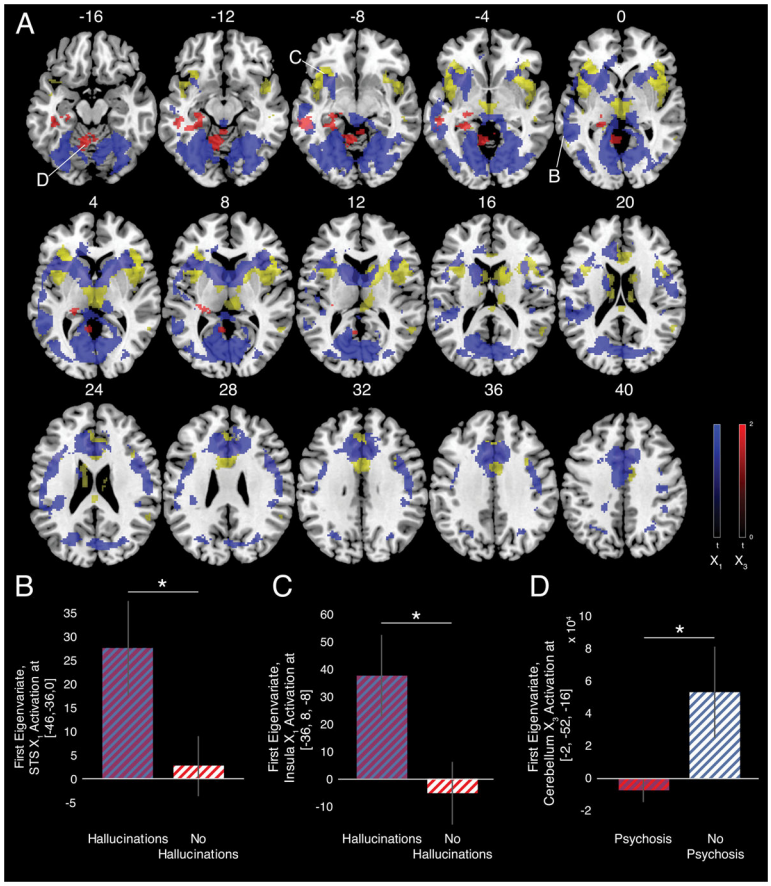

Fig 4. Hierarchical Gaussian Filter Imaging Results

A. HGF trajectories for X_1 (blue) and X_3 (red) regressed onto BOLD time courses for the conditioned hallucinations task. Regions identified significantly active during conditioned hallucinations (from Fig. 3C) are highlighted in yellow for reference. All images cluster-extent thresholded at starting value 0.05; critical k_e for X_1 = 545; X_3 = 406. B–C. Parameter estimates of X_1 fit extracted from 5-mm sphere centered on STS (B) and anterior insula (C) activation differ based upon hallucination status. D. Parameter estimates of X_3 fit extracted from 1-mm sphere centered on cerebellar vermis activation differ based upon psychosis status. Error bars represent 1 SEM.