Hallucination-Proneness is Associated With a Decrease in Robust Averaging of Perceptual Evidence

Emmett M. Larsen[1,©], Jingwen Jin[2,3], Xian Zhang[1], Kayla R. Donaldson[1], Megan Liew[1,©], Guillermo Horga[4,5], Christian Luhmann[1], and Aprajita Mohanty[1,*]

[1]Department of Psychology, Stony Brook University, Stony Brook, NY; [2]Department of Psychology, The University of Hong Kong, Hong Kong SAR, China; [3]The State Key Laboratory of Brain and Cognitive Sciences, The University of Hong Kong, Hong Kong SAR, China; [4]Department of Psychiatry, Columbia University, New York, NY; [5]New York State Psychiatric Institute (NYSPI), New York, NY

*To whom correspondence should be addressed; Department of Psychology, Stony Brook University, Stony Brook, NY, 11794-2500, USA; tel: 631.632.7800, e-mail: aprajita.mohanty@stonybrook.edu

Background and Hypothesis: Hallucinations are characterized by disturbances in perceptual decision-making about environmental stimuli. When integrating across multiple stimuli to form a perceptual decision, typical observers engage in "robust averaging" by down-weighting extreme perceptual evidence, akin to a statistician excluding outlying data. Furthermore, observers adapt to contexts with more unreliable evidence by increasing this down-weighting strategy. Here, we test the hypothesis that hallucination-prone individuals ($n = 38$ high vs $n = 91$ low) would show a decrease in this robust averaging and diminished sensitivity to changes in evidence variance. *Study Design*: We used a multielement perceptual averaging task to elicit dichotomous judgments about the "average color" (red/blue) of an array of stimuli in trials with varied strength (mean) and reliability (variance) of decision-relevant perceptual evidence. We fitted computational models to task behavior, with a focus on a log-posterior-ratio (LPR) model which integrates evidence as a function of the log odds of each perceptual option and produces a robust averaging effect. *Study Results*: Hallucination-prone individuals demonstrated less robust averaging, seeming to weigh inlying and outlying extreme or untrustworthy evidence more equally. Furthermore, the model that integrated evidence as a function of the LPR of the two perceptual options and produced robust averaging showed poorer fit for the group prone to hallucinations. Finally, the weighting strategy in hallucination-prone individuals remained insensitive to evidence variance. *Conclusions*: Our findings provide empirical support for theoretical proposals regarding evidence integration aberrations in psychosis and alterations in the perceptual systems that track statistical regularities in environmental stimuli.

Key words: perceptual averaging/psychosis/schizotypy/computational modeling/adaptive gain/liberal acceptance

Introduction

Hallucinations are characterized by disturbances of perceptual processes resulting in false decisions or inferences about the nature of environmental stimuli.[1–3] The perceptual processes involved in psychosis have often been studied so far using signal-detection paradigms that rely on detection of single target stimuli (eg,[4,5] Such paradigms, though useful, have provided limited insight, because perceptual decision-making in a complex and dynamic real-world involves not only identification of a discrete stimulus, but also integration of sensory information distributed across time or space.[6–9] One way the perceptual system deals with complexity is by extracting statistical summary representations of the mass of incoming information to form a quick perceptual decision. Anomalies in how such summary representations are formed may offer important clues into the mechanisms of hallucination formation and maintenance.

In the lab, *perceptual averaging* experiments have been used to model integration of sensory evidence across spatial arrays or temporal sequences.[8–10] This research has identified 2 key contributors to decision-making, the *strength,* and *reliability* of sensory evidence to be integrated.[10] These 2 dimensions have been described as analogous to the important considerations in statistical decision-making, in which a researcher ought to consider not only the estimates of strength (ie, mean), but also the reliability (ie, variance) of empirical evidence. A statistician comparing 2 samples of data makes a decision based on inferential statistics rather than measures of central tendency alone. Similarly, optimal perceptual

decision-making relies on effectively integrating the strength of the sensory evidence with its reliability. Computationally, it has been shown that the integration of sensory evidence when making a perceptual decision more likely reflects the ratio of the posterior probabilities of the two choice alternatives given a piece of evidence (a *log-posterior ratio* (*LPR) model*) than just the mean of the evidence (*simple averaging* model) or a *signal-to-noise ratio* obtained by scaling mean by variance (the *SNR model*).[10] Importantly, the LPR model yields "robust averaging" in which more extreme perceptual evidence is down-weighted during integration, akin to a statistician disregarding extreme outliers.[10,11] This robust decision-making strategy is thought to be adaptive in the face of varied evidence, because extreme outlying observations can lead judgments astray if allowed to exert undue influence. In line with this, greater degrees of unreliability (ie, greater variance of evidence arrays) yield increased adaptive down-weighting.[10]

Incidentally, a similar "statistician" analogy has been proposed when framing the nature of disrupted inferential processes observed in psychosis.[12] This view posited that "a key cognitive aberration in psychosis is that patients reason like 'bad statisticians', that is, that they assign meaning and momentum to weakly supported evidence" (14, p.13). In a similar vein, other work suggests that psychotic symptoms may arise in part due to a general tendency to attribute importance to irrelevant[13–16] and noisy stimuli.[17] This hypothesis can best be tested by examining whether hallucinatory symptoms are related to the aberrations in cognitive and computational processes by which stimuli varying in reliability are integrated during decision-making. In a recent novel study that examined integration of perceptual evidence in psychosis, the contribution of individual stimulus elements remained unexplored.[18]

Hence, in the present study, we sought to shed light on the mechanisms of hallucinations using a multielement perceptual averaging paradigm.[10] This allowed us to examine the influence of stimulus array elements on perceptual decisions as a function of extremeness (ie, distance from the central tendency of an array). We used a nonclinical sample, as hallucinatory experiences are measurable in attenuated form in the general population,[19] and in some cases even manifest with frequency and intensity comparable to clinical samples.[4] The influence of these experiences can be examined in such samples without the confounds associated with the diagnosis of psychotic illness (eg, effects of medication). Furthermore, preliminary work directly comparing clinical and nonclinical samples has suggested that at least some mechanisms underlying clinical hallucinations overlap with those underlying the attenuated hallucination-like experiences in the general population.[4,20] In the perceptual averaging task, we hypothesized that hallucination-prone individuals would show a decrease in the adaptive down-weighting

of extreme evidence (ie, robust averaging), a pattern that is computationally recapitulated by an LPR model. In line with this, we expected a worse fit of the LPR model for the group high in hallucination-proneness. Finally, as a secondary line of inquiry, we tested the effect of evidence variance found in past work,[10] in which robust averaging adaptively increases with evidence variance. Based on past work showing that hallucinations may relate to a diminished sensitivity to variance in environmental features,[21] we expected hallucination-proneness to be associated with a diminished sensitivity to increases in variance.

Methods

Sample and Measure of Hallucination-Proneness

Participants were undergraduate students at Stony Brook University enrolled in a psychology course. We adopted a screening procedure designed to identify individuals high and low in hallucination-proneness; this binarization has been frequently used in past work using similar measures of psychosis-proneness[22–26] and aids in obtaining clearly separated groups that provide statistical power to detect hypothesized effects. We screened a sample of 183 using the Cardiff Anomalous Perceptions Scale (CAPS; 29), a validated 32-item measure of aberrant perceptual experiences (for details regarding this scale see supplementary methods). We derived screening cutoff values for the CAPS by using the mean number of items endorsed as reported in the validation sample for the measure[27] and adding/subtracting.5 standard deviation. This resulted in a "high" group ($N = 38$) which scored 10 or more (mean CAPS = 13.8, $SD = 3.1$), and a "low" group ($N = 91$) which scored 4 or less (mean CAPS = 1.8, $SD = 1.5$). Of note, the mean number of items endorsed in our high-CAPS sample was comparable to that reported by past studies in samples of individuals with psychotic disorders,[27,28] suggesting generalizability to the clinical population.

Three participants did not complete the perceptual task due to technical difficulties, and all participants showed task accuracy above chance (50%). The final dataset thus included 126 participants, 35 high, and 91 low in hallucination-proneness (see supplementary table S1 for comparison of groups on study measures). All participants reported normal or corrected-to-normal vision and fluency in the English language. Partial course credit was awarded for study participation, and all procedures were approved by the Stony Brook IRB.

Task and Stimuli

All task and stimuli characteristics were reproduced according to specifications in de Gardelle and Summerfield[10] using the PsychoPy package for Python (version 2.7). Participants were presented with a stimulus

array composed of 8 elements arranged in a circle around a fixation point on the screen, which was positioned at an approximately 60 cm viewing distance. Each element varied in color between red and blue, according to Red-Green-Blue values: [1,0,0] and [0,0,1]. Color values were drawn randomly from Gaussian distributions parameterized with mean and standard deviation according to trial conditions, in which overall evidence strength (mean) and reliability (variance) were manipulated orthogonally. Specifically, each trial's 8 color values came from a distribution whose mean was *very red, slightly red, slightly blue,* or *very blue*, and whose variance was *high, medium,* or *low*. In analyses, we collapsed across the color factor, as it did not impact accuracy or reaction time when included in the full GLMs, and beta weights (and their quadratic trend) did not differ when estimated separately by color (all Ps > .1). The resulting design structure was thus 2 (high vs low evidence strength ["mean"]) × 3 (high, medium, or low evidence reliability ["variance"]).

Each stimulus array was presented for 2000 milliseconds; participants were required to make a binary response during this interval indicating whether they thought the array on average was "more blue" or "more red." Auditory feedback was given to indicate whether the response was correct/incorrect. See figure 1 for task structure and sample stimuli.

Computational Modeling

We investigated 3 evidence integration models taken from past work using this task.[10] Each model utilized a different approach for integrating evidence (ie, the 8 numeric color values) presented on a given trial to create a single decision variable, as follows:

- *Mean model:* Computed the arithmetic mean of the set of color values (μ).

- *SNR model:* Computed the mean and scaled this value by the variance of the set of color values (μ/σ^2).
- *LPR model:* Computed the likelihood ratio of the probabilities of each response (Red or Blue) given the array color values and took the logarithm.

Participants were fitted to each model by minimizing mean squared error between empirical and model choices. Further details of models and their implementation are presented in supplementary methods. All code used for implementing models as well as predictive checks and parameter recovery is available at https://osf.io/9vp37/.

While each of the 3 models is psychologically plausible, they differ in 2 crucial respects. First, both the SNR and LPR models are sensitive to the *variability* of evidence, while the Mean model is sensitive solely to evidence *strength.* Second, the LPR model is further distinct from the other two models in that it allows diminished contributions to its decisions from outlying compared to inlying evidence (ie, elements in an array that are far vs close to the array's mean color value). This down-weighting quality was found by de Gardelle and Summerfield[10] to roughly approximate a Gaussian pattern across element ranks and to closely reflect a weighting pattern observed in human decision-making on the task; they term this effect reproduced by the LPR model "robust averaging." Additionally, de Gardelle and Summerfield[10] noted that while the mean model is technically optimal in the present task,[29] the LPR model better approximates the robust strategy that is adaptive in conditions of ignorance about underlying generative sources of perceptual evidence.[30]

Data Analysis

Behavioral data were analyzed using SPSS version 28 except where otherwise noted. First, to quantify overall "robust averaging," we used logistic regression via the scikit-learn package for Python[31] to estimate the relative
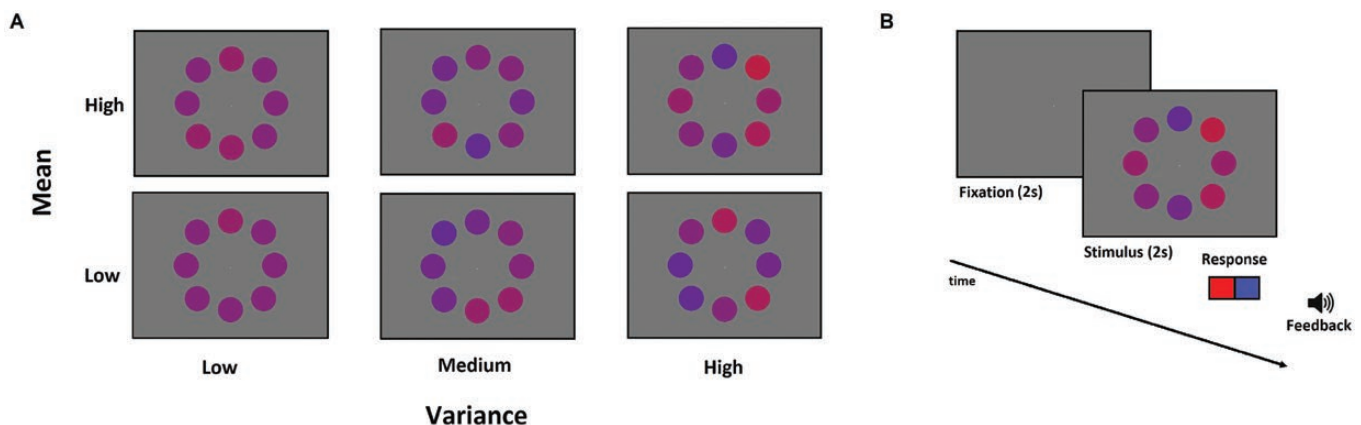

Fig. 1. Perceptual averaging task. (A) Sample stimuli from each of 6 trial-types. Only RED trials are displayed for simplicity of illustration (ie, trials in which average color was closer to red than to blue). (B) Individual trial sequence. Participants viewed 1000 stimulus arrays in total.

weights that each participant gave to array elements based on their proximity to the array mean (supplementary methods). The overall pattern of weighting the 8 elements was examined using a general linear model with element rank as the repeated measure; the magnitude of the quadratic effect provided one metric of the hypothesized "robust averaging." Lastly, we computed a second metric of robust averaging by examining the difference in average weight afforded to "inlying" (ranks 3, 4, 5, and 6) and "outlying" (ranks 1, 2, 7, and 8) elements. These analyses also included variance as a repeated-measures factor, using weights computed separately for each of the 3 variance conditions, in order to test for the hypothesized increase in adaptive down-weighting as variance increased. To test our primary hypotheses, we entered CAPS group as a between-subjects factor to test for interaction with the robust averaging metrics in the general linear models described above. Finally, we assessed the effect of variance on robust averaging in high and low-CAPS groups in order to test our hypothesis that high-CAPS would show diminished sensitivity to evidence variance.

Lastly, to examine the relation of computational indices relevant to robust averaging to measures of hallucination-proneness, we sorted participants into groups according to which of the models was best fitting for their task behavior. Then, we used Pearson's Chi-squared tests to evaluate the frequency with which high and low-CAPS participants were best fit by each model (these analyses were conducted with R programming software). As an exploratory assessment, we repeated this sorting and Chi-squared comparison separately for each variance condition.

We present several control analyses to examine robustness of results as well as specificity to hallucination-proneness over other forms of psychopathology and specificity to perceptual modality in supplementary results. We additionally include several checks related to the clinical relevance of our findings in which we construct alternative measures of hallucination-proneness which represent a tighter mapping to clinical hallucinatory manifestations. We controlled for type 1 error rate using the Benjamini–Hochberg method[32] and found that all tests reported as significant survived correction ($FDR = 0.05$).

Results

figure 2A–B displays the effects of evidence strength and reliability on choice accuracy and RT for the overall sample. These behavioral patterns replicated the initial use of the task,[10] and their statistical characterization is presented in supplementary results.

Computational Simulations of Evidence Integration

All computational models of evidence integration were able to adequately reproduce behavioral performance across conditions (see supplementary figure S5) and showed high parameter recoverability ($rs > 0.98$). In total, 55 participants were best fit by the LPR model, 62 participants were best fit by the mean model, and 9 participants were best fit by the SNR model (further information on model-fitting results is available in supplementary results).

Robust Averaging in Evidence Integration

We examined "robust averaging" across all participants by using logistic regression to estimate the relative beta weights that each participant allocated to elements based on their proximity to the array mean. figure 2C shows beta weights averaged across participants, for each element of a stimulus array, with outer values (eg, 1 and 8) representing more extreme (more blue or more red) elements that are further from the average color of the array, and moderate values (eg, 4 and 5) representing elements closer to the array average color. The distribution of weights across ranks shows an inverted-u shape (effect of element rank on the beta weights: $F_{quadratic}(1,125) = 22.61$, $P < .001$, $\eta_p^2 = 0.153$), indicating a down-weighting of the outlying evidence in the decision making. Importantly, simulations (figure 2D) showed that this behavior can be predicted by the LPR model ($F_{quadratic}(1,125) = 667.5$, $P < .001$), but not by the simple averaging model or SNR models (both $Ps > .29$). Robust averaging was not associated with any demographic variables ($Ps > .1$).

Furthermore, a significant element rank × variance interaction indicated that the magnitude of the quadratic effect differed depending on evidence variability ($F_{quad*variance}(1,125) = 6.86$, $P = .010$) such that the inverted-u shape was most pronounced for high variance (low $\eta_p^2 = 0.025$; med $\eta_p^2 = 0.061$; high $\eta_p^2 = 0.137$; figure 2E). Finally, we condensed the 8 elements into a 2-level factor consisting of inlying and outlying beta weights, to show that inlying evidence was weighted higher than outlying evidence ($F(1,125) = 25.08$, $P < .001$, $\eta_p^2 = 0.167$). As expected, this effect similarly interacted with evidence variability such that it was highest for high variance and lowest for low variance trials ($F_{invsout*variance}(2,124) = 6.34$, $P = .002$) (figure 2F).

Hallucination-Proneness and Robust Averaging

After establishing robust averaging as a strategy for evidence integration, we examined whether this strategy is modulated by hallucination proneness, measured via CAPS. First, as is depicted in figure 3A, we found that the quadratic effect of element rank differed between high ($\eta_p^2 = 0.003$) and low ($\eta_p^2 = 0.265$) CAPS groups ($F_{quad*group}(1,124) = 6.70$, $P = .011$). Next, we investigated this effect as a function of the 2-level inlying vs outlying factor; as depicted in figure 3B, the high-CAPS group showed a smaller difference between inlying and outlying weights compared to the low-CAPS group
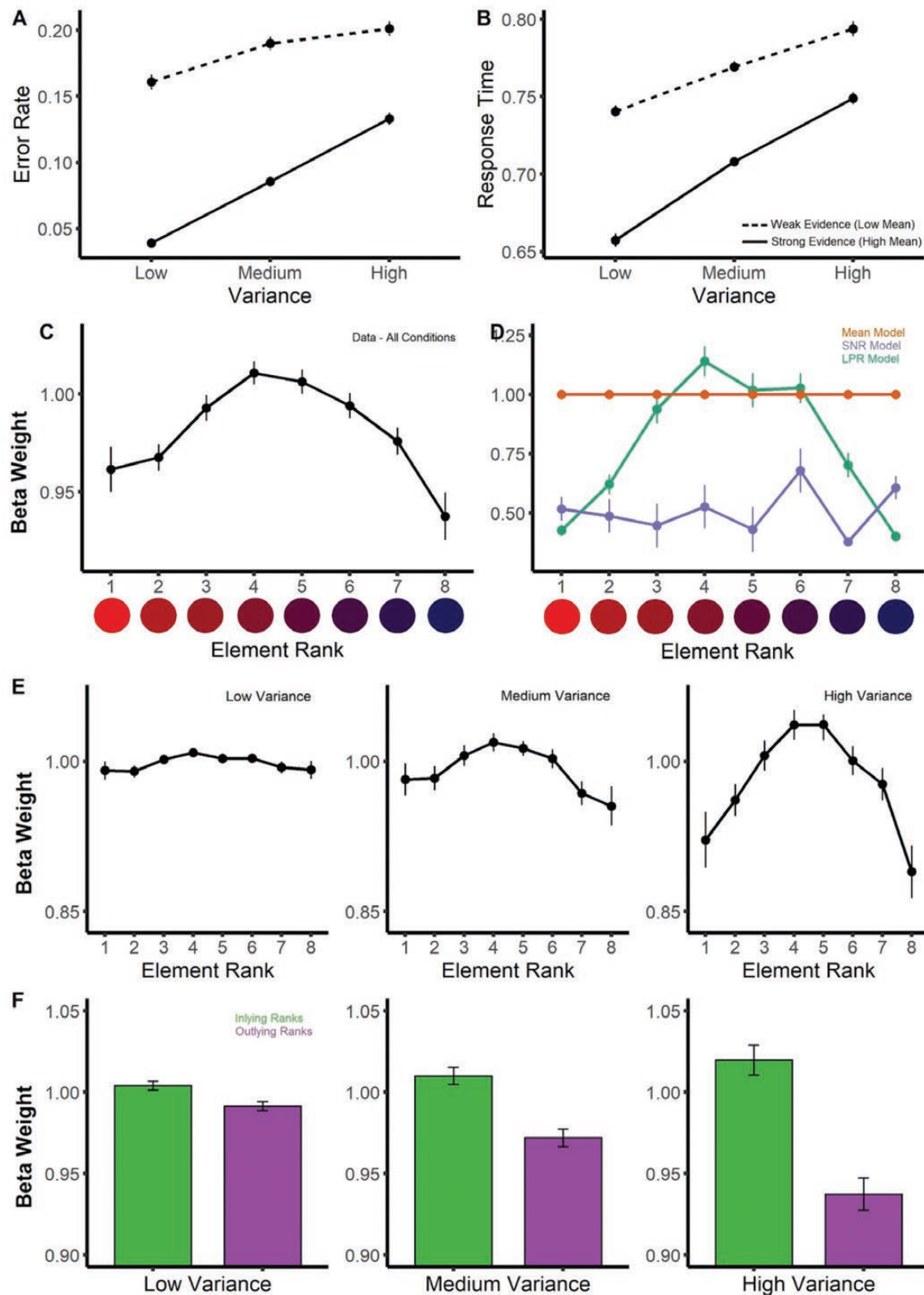
Fig. 2. Behavioral performance. (A) Error rate and (B) response times (correct trials) for weak (low mean: dotted lines) vs. strong (high mean: bold lines) evidence and for low, medium, and high variability trials (x axis). (C) Weighting of evidence across elements shown by weighting functions (estimated using logistic regression) for element ranks[1-8] with 1 and 8 being furthest from the array mean (D) Simulated weighting functions across the 8 element ranks for the mean, SNR and log-posterior-ratio models. (E) Weighting of evidence (empirical) for low, medium, and high-variance trials (F) Weighting patterns visualized as 2-level factor (inlying and outlying evidence weights) for low, medium, and high-variance trials.
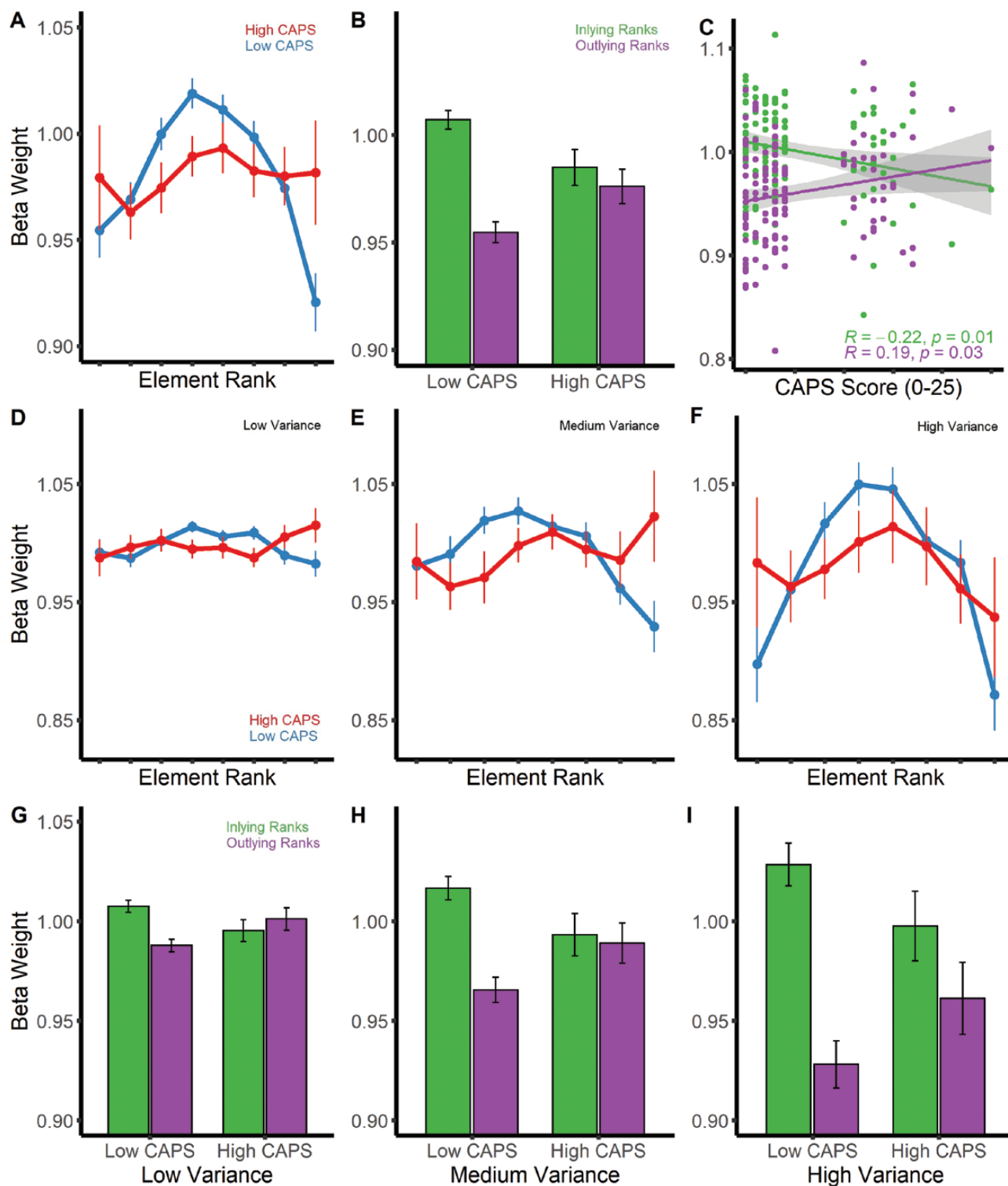
Fig. 3. Influence of hallucination-proneness on weighting of evidence. (A) Weighting of evidence across elements for high vs low hallucination-prone groups. (B)-(C) Weighting of inlying and outlying evidence as a function of hallucination-proneness, depicted for (B) high vs low hallucination-prone groups, and (C) continuously (Pearson's correlations indicated on the figure). (D)-(F) Weighting of evidence across elements by high vs low hallucination-prone groups when variability of evidence is (D) low (E) medium and (F) high. (G)-(I) Differential weighting of inlying vs outlying evidence by high vs low hallucination-prone groups for (G) low, (H) medium, and (I) high evidence variability.

($F_{\text{invsout*group}}(1,124) = 6.10$, $P = .015$; also see supplementary figure S11 for this figure presented as paired boxplots). figure 3C depicts the continuous version of this effect, showing that as CAPS increased, inlying evidence was weighted less, and outlying evidence was weighted more. (These effects can also be viewed separately for all 6 conditions in (supplementary figures S6–S7.)

Lastly, to test the responsivity of weighting patterns to evidence variance, we examined the effect of hallucination proneness on robust averaging for low, medium, and high evidence variability. As seen in figure 3D–F, the distribution of weights across element ranks for the low-CAPS group shows an increasing inverted-u shape (quadratic effect) as evidence variability increased ($F_{\text{quad*variance}}(1,90) = 5.02$, $P = .028$) such that the quadratic effect increased from low ($\eta_p^2 = 0.081$), to medium ($\eta_p^2 = 0.137$) to high ($\eta_p^2 = 0.192$), but this effect was not observed for the high-CAPS group ($F_{\text{quad*variance}}(1,34) = 1.79$, $P = .19$; low $\eta_p^2 = 0.020$, medium $\eta_p^2 = 0.003$, high $\eta_p^2 = 0.032$). In line with this, figure 3G–I shows that for the low-CAPS group, down-weighting of outlying evidence compared to inlying evidence increased as evidence variability increased ($F_{\text{invsout*variance}}(2,89) = 5.72$, $P = .005$, $\eta_p^2 = 0.114$), but this pattern was not observed for the high-CAPS group ($F_{\text{invsout*variance}}(2,33) = 0.985$, $P = .384$, $\eta_p^2 = 0.056$).

Effect of Hallucination-Proneness on Evidence Integration

Finally, we examined the relationship between computational indices of evidence integration and psychosis proneness. We divided the sample into groups based on whether they were best fit by the Mean or LPR model, and compared group membership to membership in high vs low hallucination-proneness group. (due to the small number of participants best-fit by the SNR model, we excluded it from these analyses; of the 9 participants best-fit by this model, 5 were low-CAPS, and 4 were high-CAPS.) Pearson's Chi-squared tests revealed that overall a greater number of high-CAPS compared to low-CAPS participants were best-fit by the Mean model vs the LPR model ($\chi^2(1) = 5.47$, $P = .019$) (figure 4A). We recomputed the categorization of participants by their best-fitting models separately by variance (figure 4B-D) to explore whether this grouping pattern differed by variance condition, and we observed that the difference was significant in the high-variance condition only ($\chi^2(1) = 4.27$, $P = .039$) (figure 4D). (Versions of these analyses conducted using the continuous CAPS score can also be viewed in (supplementary figure S8).

Discussion

While hallucinations are thought to involve disordered perceptual processes, little work has examined the role of differential evidence characteristics during dynamic gathering and integrating of perceptual information. Here, we used a multielement averaging task in which we manipulated the strength and reliability of the decision-relevant perceptual information. We sought to test whether hallucination-proneness is related to altered weighting of perceptual evidence and diminished responsivity to evidence variance. We assessed these questions in a nonclinical sample psychometrically prone to hallucination-like experiences in order to isolate relevant mechanisms without concern for confounding effects of psychotic illness and its sequelae.

In line with past work,[10] we found that observers overall tended to engage in robust averaging (ie, down-weighting more extreme (outlying) elements of the array) when forming a summary perceptual decision, analogous to a statistician attuned to the potentially lower trustworthiness of outliers. Of note, when evaluating computational models in our sample, a sizable number of participants across all groups were better fit by the Mean model (rather than the LPR), suggesting that an important direction for future work is to refine models of the processes leading to the robust averaging observed in participant behavior.

A novel finding of the present study is that those high in hallucination-proneness demonstrated less of the robust averaging strategy, seeming to weigh inlying, and outlying evidence more equally. Furthermore, the model that integrated evidence as a function of the LPR of the two perceptual options and produced robust averaging showed poorer fit for the group prone to hallucinations. Our findings also showed that hallucination- but not delusion-proneness (see supplementary results) is associated with alterations in the perceptual systems involved in integrating disparate sensory evidence.

The present findings provide empirical support for theoretical proposals regarding evidence integration aberrations in psychosis. Moritz et al[12] proposed that evidence integration in psychosis is characteristically un-statistician-like, such that positive symptoms are thought to be associated with the tendency to attribute "meaning and momentum to weakly supported evidence (14, p.13). This pattern has been termed "liberal acceptance,"[12] which might be aptly used to describe the non-robust weighting strategy we observed in our hallucination-prone group. Similarly, the aberrant salience framework has proposed that in psychosis, unimportant stimuli are experienced as imbued with strong decisional weight.[15,16] These patterns may also provide insight into empirical reports suggesting that individuals with psychosis imbue noisy unreliable information with meaning.[17]

We further observed that hallucination-proneness was associated with insensitivity to evidence variance: while low-hallucination-prone observers were responsive to increased variance by employing an increased down-weighting strategy, high-hallucination-prone observers
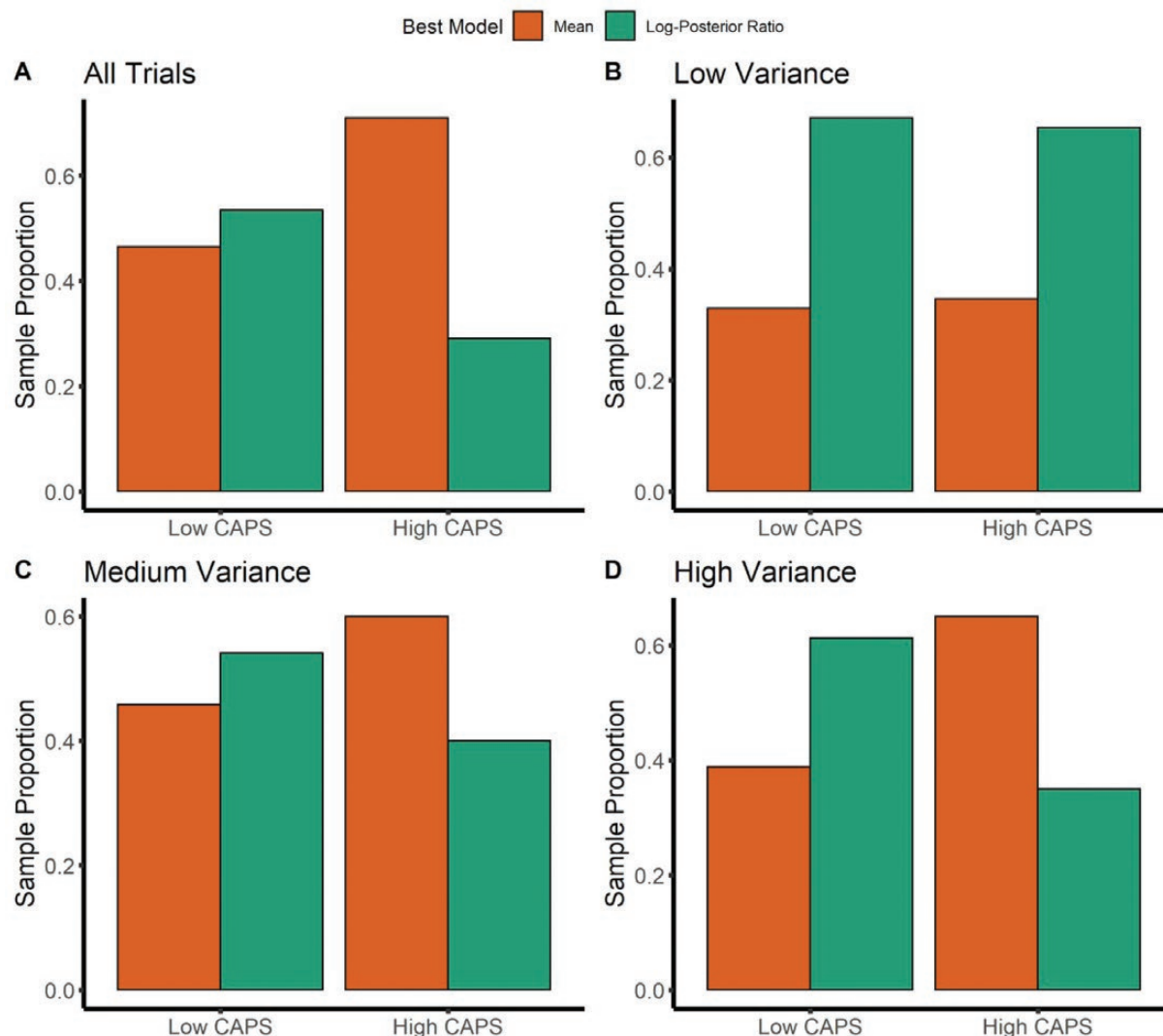
Fig. 4. Proportion of participants in each Cardiff Anomalous Perceptions Scale group that were best fit by the Mean model and the log-posterior-ratio model when fits estimated (A) from all trials, and (B-D) separately by variance condition.

did not make this adjustment. We also found that the high-variance condition was best able to discriminate CAPS groups in terms of model fits, likely resulting from the strongest tendency in this condition for non-hallucination-prone observers to use the robust strategy. These findings are consistent with past work by Cassidy et al,[21] who found a similar insensitivity to variance of environmental stimuli in patients with more severe hallucinations. They attributed this effect to the presence of "strong priors" for those with higher hallucinations, which resulted in aberrant performance particularly in high-variance contexts when priors should be down-weighted. (Of note, our model-fitting results may also be consistent with the importance of high-variance contexts, as this condition seemed to show the greatest separation between high- and low-CAPS groups.) Their study further showed evidence for a dopaminergic substrate of this effect in the associative striatum as well as a relationship

with decreased gray matter volume in the dorsal ACC; these candidate neurobiological mechanisms could be probed in future work to assess whether they also subserve the variance-insensitivity we observed in the present work.

Future Directions

Future work should attempt integration of the present perceptual averaging research with Bayesian approaches to modeling evidence integration processes in psychosis.[1] Work in this vein has adopted the predictive coding framework and found evidence linking psychosis and positive symptoms in particular to aberrant priors.[3,21,33,34] Evidence is amassing that abnormalities are present in psychosis during the process of integrating prior, top–down expectations with bottom–up sensory information to reach perceptual inferences. While some work suggests

that in low-level perceptual tasks, individuals with psychosis show *weak priors*, as evidenced, for example, by a decreased tendency to experience visual illusions,[35,36] other work suggests a relationship between psychosis with *stronger priors* at higher levels of information processing (see 36 for review). The present paradigm unfortunately does not permit straightforward interpretations in terms of the use of priors, as it does not involve ongoing updating of perceptual beliefs. However, future modification of our paradigm could extend this line of work by delineating effects of evidence characteristics (eg, extremeness of each component) while simultaneously examining individual variation in reliance on priors. Further, neuroimaging studies should capitalize on the simple perceptual task implemented here in order to identify neural substrates of these hallucination-related effects. Potential correlates may include midbrain dopamine transmission regions (such as those implicated in 23), and auditory cortex, which may play a role in altered prediction error signaling associated with hallucinations.[37] Basic work has also shown that the parietal cortex contains functional correlates which track the log-likelihood ratio during perceptual evidence integration.[38]

In addition, it is crucial that future work attempt to replicate these findings in a clinical sample in order to confirm their utility in relation to functional outcomes in psychotic illness. It also remains critically important for mechanistic work like ours to be validated through tests in clinical samples for other reasons. Namely, while some work suggests overlap in the mechanisms of clinical phenomena and their attenuated analogs in the nonclinical population, the same studies also call for caution in that area of non-overlap are also evident.[20] Moreover, the most direct tests of shared/unshared mechanisms of hallucinations have been conducted in samples of psychics and mediums with no psychotic disorder diagnosis (eg, 4, 22) who may also differ in other unmeasured ways from general population samples. Thus it remains an important unanswered empirical question to distinguish which mechanisms are and are not shared between these populations, and the present results should be considered preliminary until such tests are conducted. Of note, several robustness-check analyses (presented in our supplementary results-clinical relevance) provide some reason to be optimistic about the replicability of our findings in samples with clinical psychosis. However, other supplemental analyses (see supplementary results-specificity checks-CAPS factors) may suggest that our primary effects relate most strongly to *perceptual distortion* experiences, whose mechanisms may differ from those underlying clinical hallucinations. Finally, it is of note that clinical hallucinations occur most commonly in the auditory modality (eg,[39]), and as such future work could adapt the present paradigm to examine replicability in that and other sensory dimensions.

References

1. Adams RA, Stephan KE, Brown HR, Frith CD, Friston KJ. The computational anatomy of psychosis. *Front Psychiatry.* 2013;4:47. doi: 10.3389/fpsyt.2013.00047.
2. Benrimoh D, Parr T, Vincent P, Adams RA, Friston K. Active inference and auditory hallucinations. *Comput Psychiatry.* 2018;2:183–204.
3. Horga G, Abi-Dargham A. An integrative framework for perceptual disturbances in psychosis. *Nat Rev Neurosci.* 2019;20(12):763–778.
4. Powers AR, Mathys C, Corlett PR. Pavlovian conditioning–induced hallucinations result from overweighting of perceptual priors. *Science.* 2017;357(6351):596–600. doi: 10.1126/science.aan3458
5. Vercammen A, de Haan EHF, Aleman A. Hearing a voice in the noise: auditory hallucinations and speech perception. *Psychol Med.* 2008;38(8):1177–1184. doi: 10.1017/S0033291707002437
6. Heekeren HR, Marrett S, Ungerleider LG. The neural systems that mediate human perceptual decision making. *Nat Rev Neurosci.* 2008;9(6):467–479.
7. Gold JI, Shadlen MN. The neural basis of decision making. *Annu Rev Neurosci.* 2007;30:535–574.
8. Albrecht AR, Scholl BJ. Perceptually averaging in a continuous visual world: extracting statistical summary representations over time. *Psychol Sci.* 2010;21(4):560–567.
9. Whitney D, Yamanashi Leib A. Ensemble perception. *Annu Rev Psychol.* 2018;69:105–129.
10. de Gardelle V, Summerfield C. Robust averaging during perceptual judgment. *Proc Natl Acad Sci U S A.* 2011;108(32):13341–13346. doi: 10.1073/pnas.1104517108
11. Rousseeuw, P. J., Leroy, A. M. *Robust regression and outlier detection. Wiley series in probability and mathematical statistics.* United States, John Wiley & Sons, Inc.; 1987.
12. Moritz S, Pfuhl G, Lüdtke T, Menon M, Balzan RP, Andreou C. A two-stage cognitive theory of the positive symptoms of psychosis. Highlighting the role of lowered decision

thresholds. *J Behav Ther Exp Psychiatry.* 2017;56:12–20. doi: 10.1016/j.jbtep.2016.07.004

13. Kapur S. Psychosis as a state of aberrant salience: a framework linking biology, phenomenology, and pharmacology in schizophrenia. *Am J Psychiatry.* 2003;160(1):13–23. doi: 10.1176/appi.ajp.160.1.13

14. Howes OD, Hird EJ, Adams RA, Corlett PR, McGuire P. Aberrant salience, information processing, and dopaminergic signaling in people at clinical high risk for psychosis. *Biol Psychiatry.* 2020;88(4):304–314. doi: 10.1016/j.biopsych.2020.03.012

15. Howes OD, Nour MM. Dopamine and the aberrant salience hypothesis of schizophrenia. *World Psychiatry.* 2016;15(1):3–4. doi: 10.1002/wps.20276

16. Broyd A, Balzan RP, Woodward TS, Allen P. Dopamine, cognitive biases and assessment of certainty: a neurocognitive model of delusions. *Clin Psychol Rev.* 2017;54:96–106. doi: 10.1016/j.cpr.2017.04.006

17. Galdos M, Simons C, Fernandez-Rivas A, *et al*. Affectively salient meaning in random noise: a task sensitive to psychosis liability. *Schizophr Bull.* 2011;37(6):1179–1186. doi: 10.1093/schbul/sbq029

18. Bansal S, Bae GY, Robinson BM, *et al*. Association between failures in perceptual updating and the severity of psychosis in schizophrenia. *JAMA Psychiatry.* 2022;79(2):169–177.

19. Johns LC, van Os J. The continuity of psychotic experiences in the general population. *Clin Psychol Rev.* 2001;21(8):1125–1141. doi: 10.1016/S0272-7358(01)00103-9

20. Moseley P, Alderson-Day B, Common S, *et al*. Continuities and discontinuities in the cognitive mechanisms associated with clinical and nonclinical auditory verbal hallucinations. *Clin Psychol Sci.* 2022;10(4): 725–766. doi:10.1177/21677026211059802

21. Cassidy CM, Balsam PD, Weinstein JJ, *et al*. A perceptual inference mechanism for hallucinations linked to striatal dopamine. *Curr Biol.* 2018;28(4):503–514.e4. doi: 10.1016/j.cub.2017.12.059

22. Anandakumar T, Connaughton E, Coltheart M, Langdon R. Belief-bias reasoning in non-clinical delusion-prone individuals. *J Behav Ther Exp Psychiatry.* 2017;56:71–78. doi: 10.1016/j.jbtep.2017.02.005

23. Ashinoff BK, Buck J, Woodford M, Horga G. The effects of base rate neglect on sequential belief updating and real-world beliefs. *PLoS Comput Biol.* 2022;18(12):e1010796. doi: 10.1371/journal.pcbi.1010796

24. Diaconescu AO, Wellstein KV, Kasper L, Mathys C, Stephan KE. Hierarchical Bayesian models of social inference for probing persecutory delusional ideation. *J Abnorm Psychol.* 2020;129:556–569. doi: 10.1037/abn0000500

25. Na S, Blackmore S, Chung D, *et al*. Computational mechanisms underlying illusion of control in delusional individuals. *Schizophr Res.* 2022;245:50–58. doi: 10.1016/j.schres.2022.01.054

26. Wellstein KV, Diaconescu AO, Bischof M, *et al*. Inflexible social inference in individuals with subclinical persecutory delusional tendencies. *Schizophr Res.* 2020;215:344–351. doi: 10.1016/j.schres.2019.08.031.

27. Bell V, Halligan PW, Ellis HD. The Cardiff Anomalous Perceptions Scale (CAPS): a new validated measure of anomalous perceptual experience. *Schizophr Bull.* 2006;32(2):366–377. doi: 10.1093/schbul/sbj014

28. Bell V, Halligan PW, Pugh K, Freeman D. Correlates of perceptual distortions in clinical and non-clinical populations using the Cardiff Anomalous Perceptions Scale (CAPS): associations with anxiety and depression and a re-validation using a representative population sample. *Psychiatry Res.* 2011;189(3):451–457. doi: 10.1016/j.psychres.2011.05.025

29. van den Berg R, Ma WJ. Robust averaging during perceptual judgment is not optimal. *Proc Natl Acad Sci U S A.* 2012;109(13):E736; author reply R737–736; author reply E736. doi: 10.1073/pnas.1119078109

30. de Gardelle V, Summerfield C. Reply to van den Berg and Ma: robust decision makers are not omniscient. *Proc Natl Acad Sci U S A.* 2012;109(13):E737–E737. doi: 10.1073/pnas.1120640109

31. Pedregosa F, Varoquaux G, Gramfort A, *et al*. Scikit-learn: machine learning in python. *J Mach Learn Res.* n.d.;12:2825–2830

32. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Series B Stat Methodol.* 1995;57(1):289–300.

33. Corlett PR, Horga G, Fletcher PC, Alderson-Day B, Schmack K, Powers AR. Hallucinations and strong priors. *Trends Cogn Sci.* 2019;23(2):114–127. doi: 10.1016/j.tics.2018.12.001

34. Sterzer P, Voss M, Schlagenhauf F, Heinz A. Decision-making in schizophrenia: a predictive-coding perspective. *Neuroimage.* 2019;190:133–143. doi: 10.1016/j.neuroimage.2018.05.074

35. Dima D, Roiser JP, Dietrich DE, *et al*. Understanding why patients with schizophrenia do not perceive the hollow-mask illusion using dynamic causal modelling. *Neuroimage.* 2009;46(4):1180–1186. doi: 10.1016/j.neuroimage.2009.03.033

36. Keane BP, Silverstein SM, Wang Y, Papathomas TV. Reduced depth inversion illusions in schizophrenia are state-specific and occur for multiple object types and viewing conditions. *J Abnorm Psychol.* 2013;122(2):506–512. doi: 10.1037/a0032110

37. Horga G, Schatz KC, Abi-Dargham A, Peterson BS. Deficits in predictive coding underlie hallucinations in schizophrenia. *J Neurosci.* 2014;34(24):8072–8082. doi: 10.1523/JNEUROSCI.0200-14.2014

38. Yang T, Shadlen MN. Probabilistic reasoning by neurons. *Nature.* 2007;447(7148):1075–1080.

39. McCarthy-Jones, S, Smailes, D, Corvin, A, *et al*. Occurrence and co-occurrence of hallucinations by modality in schizophrenia-spectrum disorders. *Psychiatry Res.* 2017;252:154–160.