The changing weight of expectation: how shifting priors underpin variability in hallucination frequency.

Commentary on Kafadar et al: Conditioned hallucinations and prior over-weighting are state sensitive markers of hallucination susceptibility

Fletcher PC[1,2*] and Teufel CR[3]

[1] Department of Psychiatry, University of Cambridge, Herchel Smith Building, Addenbrooke's Hospital, Cambridge CB2 0SZ, UK and Cambridgeshire and Peterborough NHS Trust, Cambridge, UK

[2] Wellcome Trust MRC Institute of Metabolic Science, University of Cambridge, Cambridge Biomedical Campus, Cambridge CB2 0QQ, UK

[3] School of Psychology Cardiff University Brain Research Imaging Centre, Maindy Road, Cardiff, CF24 4HQ

* Corresponding Author: Paul Fletcher pcf22@cam.ac.uk

The predictive processing framework encourages us to consider perception as arising from the integration of pre-existing knowledge with current sensory input (**1**). It formulates hallucinations - percepts without accompanying stimuli - in terms of increased weighting of predicted over actual input and thereby suggests an explanation of the computational mechanisms underlying such experiences. Several studies of individuals with, or prone to, hallucinations have exploited this framework and have suggested ways in which the balance between prediction and input may be altered. In this issue, Kafadar and colleagues (**2**) elegantly extend this work in three ways: first, they replicate a previous observation (**3**) that individuals experiencing auditory-verbal hallucinations show a greater influence of prediction on auditory experience; second, they show that this tendency reflects individuals' current levels of auditory hallucinations rather than a persisting trait. Third, they validate an online version of their task, which may prove convenient in its future use.

Participants were grouped, using a self-report screening tool, into those with auditory hallucinations (AH+) and those without (AH-). Having learned a reliable association between a visual stimulus and an auditory tone, over the course of the experiment, they systematically reduced the proportion of trials in which the visual stimulus predicted the tone. The pre-learned expectancies led to "conditioned hallucinations" (CHs), i.e. subjective reports of hearing tones when there were none present, an effect that was significantly greater in the AH+ group. Two findings suggest that CHs track the participant's current state. First, in a sub-group of participants, detailed phenomenological analysis showed that CH-rate was related to levels of AHs two days prior to testing rather than to more remote measures at the "worst time in their history". Second, in a smaller group tested a second time, changes in frequency of AHs were associated with changes in the rate of CHs. Computational modelling of the data suggested that a parameter signalling the relative weighting of the "prior" about probability of a tone (given a visual stimulus) captures this effect.

As well as replicating an important observation (**3**) in a larger, heterogeneous sample this study provides a demonstration that the task is effective when run online: an impressive achievement with implications for its use in ever-larger samples. The observation that the behavioural phenomenon (CH) and the putative underlying process (overweighting of the prior) reflect a state rather than a trait measure has important implications both for our understanding of the mechanism and, potentially, for the value of the task as a clinical tool in monitoring symptoms and assessing therapeutic interventions.

Of course, it is well-known (**4**) that hallucinations, notably in schizophrenia, are associated with a tendency to hear what one has been led to expect, or at least to report hearing it, which is not, of course, the same thing. It is hard to be sure that responding "yes" in a sound-absent trial truly reflects a perceptual experience rather than an alteration in response or decision biases. For example, even in the absence of a genuine change in perception, participants might show an increased tendency to affirm that they have heard a sound simply based on a cognitive assessment that a sound is probable. Distinguishing genuine perceptions from non-perceptual biases is exceedingly difficult and has driven complex and sophisticated methodologies in psychophysics. While the previous study (**3**) ingeniously used BOLD signal in auditory cortex as a marker for perception, this too is not without significant ambiguity. As ever, such ambiguities will likely be ironed out using targeted experimental designs in future studies. Interestingly, a comparable study examined expectancy-driven visual hallucinations in schizophrenia (**5**). Using a dot motion task, Valton and colleagues examined the impact of prior learning on both dot detection and direction-estimation. The impact of prior expectations in this task was found to manifest in two ways: an increased tendency to perceive dots moving in the most expected directions and a tendency to perceive dots moving in those directions even when no dots were present. However, this effect was weaker in people with schizophrenia than controls, being more in accord with previous observations of "weakened perceptual priors" (**6**). It is possible that

priors in the context of psychosis may prove either weak or strong depending on the nature or level of the prediction (**6**). In the visual domain, for example, predictions relating to semantic content are demonstrably stronger in psychosis-prone individuals (**7**) and, moreover, priors expressed at different levels are differentially sensitive to separate psychosis-like features. Employing comparable concepts within a different framework, Jardri and colleagues showed that people with schizophrenia tend to overweight sensory evidence rather than prior information (**8**).

However, the above studies were in the visual domain and evidence from studies of auditory processing accords more strongly with the idea that excessive weighting of expectations is a feature of hallucinations. Cassidy and colleagues, using an auditory duration-estimation task not only implicated excessive weighting of priors, as a consequence, they argued, of over-estimation of the precision of these priors, but also provided compelling evidence linking this to dopamine dysregulation (**9**). This fits well with the current data. Intriguingly, in experiments examining speech detection (**10**), hallucination-proneness was not associated with an enhanced “top-down” tendency when participants were looking for speech within a noisy auditory signal but there was a pre-existing tendency to expect, and successfully decipher, speech. Taken together with the current findings, it appears that AVHs may be associated both with a state-dependent fluctuation in prior weighting and a more persistent tendency to predict speech in auditory signal.

In addition to the behavioural analysis, Kafadar and colleagues applied a computational model to elucidate mechanisms underlying their observations. This analysis identifies a key parameter distinguishing the AH+ and AH- groups, taken to signify the relative contribution of the prior to the perceptual process. This is the ratio of the precision of the prior (expectation that there will be a tone given the presence of the visual stimulus) to the precision of the sensory experience. Again, some ambiguity should be acknowledged, partly given the difficulty, as discussed, in distinguishing a genuine percept from a decision/response bias, as discussed. In addition, there is more than one way, mechanistically, in which prediction could exert its effects on perception (1). Nonetheless, future studies may resolve these ambiguities and this (2, 3) and related work provides an excellent foundation for beginning to isolate the mechanistic underpinnings. Meanwhile, the observation that the weighting of the prior parameter relates to AHs in a state-dependent way offers grounds for speculating. It seems unlikely that this marker reflects a structural alteration, which would presumably be less likely to show such fluctuations. Rather, it may signal a dynamic mechanism perhaps underpinned by fluctuating modulatory neutrotransmitter activity (2). Intriguingly, a closer look at the longitudinal data suggests that the link between changes in CH tendency and changes in AH frequency appears to be driven by a surprising effect (see supplemental figure 6). The only individuals showing an increase in AH frequency were those who had previously reported no experience of hallucinations. Indeed, judging by figure 4B, groups experiencing a decrease, or no change, in AHs showed little or no alteration in CH rate. It is unclear whether this transition in some participants reflects the emergence of true hallucinations or of non-specific unusual auditory experiences but its occurrence, and its accompaniment by an increased rate of CHs, is noteworthy.

Turning to the potential clinical value, the study nicely exemplifies the aspirations of computational psychiatry. The authors offer an analogy: monitoring levels of Thyroid Stimulating Hormone (TSH) serves as a practically useful biomarker for the functional integrity of the thyroid gland. Crucially, the symptoms of reduced thyroid gland function (fatigue, depression, aches, weight gain) are neither causally driven by excessive TSH (which is a reaction to the dysfunction) nor are they always present even when TSH is abnormal. The TSH, in this instance, may be seen as the canary in the coalmine, a marker that may require action (though this action would not be directed at the canary itself). While instructive, the analogy perhaps understates the implications of the current study findings

Like TSH, CHs and stronger priors may mark the presence of a process that could have negative consequences but, unlike, TSH, they are seen as a core part of this process and their modification could have therapeutic benefits. We should be cautious however: such a link is extremely interesting but further work will be required to determine its clinical value. While the authors rightly suggest that, as a marker, CHs could serve a similar function to TSH, but it would be crucial to demonstrate superiority over alternative measure of AHs. Does the measurement of CHs yield information guiding clinical management that existing, briefer and simpler data-gathering exercises (such as a clinical interview or self-report) do not? This is an key question determining clinical value. Notwithstanding this, however, it is undeniable that the current study offers us useful insights to the mechanisms underlying AHs and, even if clinical utility proves more limited than the authors suggest, this is a valuable contribution.

Financial Disclosures

PCF has received consultancy fees from Ninja Theory Ltd, UK and from Glaxo SmithKline and is on the advisory board for Hooke (UK)

CRF reports no financial interests.

References

1: Teufel C, Fletcher PC. Forms of prediction in the nervous system. Nat Rev Neurosci. 2020 Apr;21(4):231-242.

2. Eren Kafadar *et al.* (in press): Conditioned hallucinations and prior over-weighting are state-sensitive markers of hallucination susceptibility. *Biol Psychiatry* XX:XXX-XXX

3. Powers AR, Mathys C, Corlett PR. Pavlovian conditioning-induced hallucinations result from overweighting of perceptual priors. Science. 2017 Aug 11;357(6351):596-600.

4. Mintz S, Alpert M. Imagery vividness, reality testing, and schizophrenic hallucinations. J Abnorm Psychol. 1972 Jun;79(3):310-6.

5. Valton V, Karvelis P, Richards KL, Seitz AR, Lawrie SM, Seriès P. Acquisition of visual priors and induced hallucinations in chronic schizophrenia. Brain. 2019 Aug 1;142(8):2523-2537.

6. Schmack K, Gòmez-Carrillo de Castro A, Rothkirch M, Sekutowicz M, Rössler H, Haynes JD, Heinz A, Petrovic P, Sterzer P. Delusions and the role of beliefs in perceptual inference. J Neurosci. 2013 Aug 21;33(34):13701-12.

7. Teufel C, Subramaniam N, Dobler V, Perez J, Finnemann J, Mehta PR, Goodyer IM, Fletcher PC. Shift toward prior knowledge confers a perceptual advantage in early psychosis and psychosis-prone healthy individuals. Proc Natl Acad Sci U S A. 2015 Oct 27;112(43):13401-6.

8. Jardri R, Duverne S, Litvinova AS, Denève S. Experimental evidence for circular inference in schizophrenia. Nat Commun. 2017 Jan 31;8:14218.

9. Cassidy CM, Balsam PD, Weinstein JJ, Rosengard RJ, Slifstein M, Daw ND, Abi-Dargham A, Horga G. A Perceptual Inference Mechanism for Hallucinations Linked to Striatal Dopamine. Curr Biol. 2018 Feb 19;28(4):503-514.e4.

10. Alderson-Day B, Moffatt J, Lima CF, Krishnan S, Fernyhough C, Scott SK, Denton S, Leong IYT, Oncel AD, Wu YL, Gurbuz Z, Evans S. Susceptibility to auditory hallucinations is associated with spontaneous but not directed modulation of top-down expectations for speech. Neurosci Conscious. 2022 Feb 1;2022(1):niac002.

Acknowledgements:
Funding was provided by the Bernard Wolfe Health Neuroscience Fund and a Wellcome Trust Investigator Award to PCF (Reference No. 206368/Z/17/Z). This work was supported by the NIHR Cambridge Biomedical Research Centre (BRC-1215-20014).