Title

Perceptual expectations and false percepts generate stimulus-specific activity in distinct layers of the early visual cortex

Joost Haarsma[1], Narin Deveci[1], Nadège Corbin[1,2], Martina F. Callaghan[1], Peter Kok[1]
[1] Wellcome Centre for Human Neuroimaging, UCL Queen Square Institute of Neurology, University College London, London, UK. [2] Centre de Résonance Magnétique des Systèmes Biologiques, UMR5536, CNRS/University Bordeaux, Bordeaux, France

Abstract

Popular theories suggest that hallucinations arise through excessive top-down perceptual expectations, whereas others have emphasised the role of spontaneous bottom-up activity. These theories make different predictions about how input and feedback layers in sensory regions contribute to hallucinations. Here, we used layer-specific fMRI to interrogate neural activity underlying hallucinations – high confidence false percepts – and perceptual expectations while healthy participants (N=25) performed a perceptual discrimination task. We found that false percepts were related to stimulus-like activity in the middle input layers of V2. On the other hand, perceptual expectations activated the deep feedback layers of V2, without influencing perception. The prevalence of high confidence false percepts was related to everyday hallucination severity, confirming their ecological validity. These results reveal that hallucinations can arise through spontaneous stimulus-specific activity in the input layers of the visual cortex, independent of top-down perceptual expectations.

Introduction

Hallucinations – vivid percepts in the absence of corresponding sensory inputs – are a key symptom of various psychiatric and neurological disorders. However, the underlying neural mechanisms remain hotly debated.

The previous decade has seen a rise in the popularity of top-down models of hallucinations, which propose that hallucinations arise through an excessive influence of top-down perceptual expectations[1–4]. One particularly dominant model is the predictive coding theory of hallucinations[2]. According to predictive processing theories, the brain models the world by continuously deploying perceptual expectations regarding the most likely stimuli[5,6]. These perceptual expectations are subsequently integrated with sensory inputs to form a percept. The more strongly expectations are weighted, the more they will influence perception, driving it away from sensory input. Therefore, a hallucination might arise when perceptual expectations become excessively overweighted[1–4,7]. Indeed, there are many studies demonstrating that expectations play an integral role in shaping sensory experience[8]. Furthermore, there is increasing evidence that individuals experiencing hallucinations demonstrate a stronger influence of perceptual expectations on perception[9–15]. Due to the central role of internally generated top-down perceptual expectations, these models predict that hallucinated percepts would be reflected in the layers of sensory cortex that convey feedback signals.

However, hallucinations are also known to arise from fluctuations in feedforward sensory signals[16,17]. For example, in Charles Bonnet syndrome, the visual cortex is partially deafferented, which causes it to become hyperexcitable[4,17–19]. Recent studies of this syndrome suggest that spontaneous local activity emerges in early visual cortex and is fed up the cortical hierarchy, eventually generating a hallucinatory percept[16,19]. This demonstrates that hallucinations can at least in principle emerge from feedforward processes. Further, some studies in psychosis have found increased reliance on sensory evidence, particularly with relation to delusions[9,20–25]. Additionally, there is evidence that spontaneous activity can contribute to the experience of false percepts in healthy individuals as well, although it is at present unclear whether these reflect feedforward or feedback signals[26–30]. If hallucinations do emerge through feedforward signal fluctuations, they would be expected to be reflected in the input layers of the sensory cortex.

Sensory feedforward and feedback signals are separated on a laminar level, as they preferentially terminate in different cortical layers. That is, cortical feedback neurons terminate in the agranular deep (layers V and VI) and superficial layers (layers I, II and III), whereas feedforward input preferentially terminates in the middle granular layers (layer IV) of the visual cortex[31,32]. It has recently become possible to non-invasively measure the activity in different cortical layers in humans using layer-specific fMRI, allowing testing of theories about the contributions of feedforward and feedback signals in cognition and perception[7,33–37].

In the present study, feedforward and feedback hypotheses of hallucinations were distinguished using layer-specific fMRI in healthy human participants. If hallucinations are driven by top-down feedback, false percepts should be reflected by activity in the deep layers of the early visual regions, similar to perceptual expectations[38]. By contrast, if hallucinations are driven by spontaneous feedforward activity, false percepts are expected to be reflected in the middle input layers of the visual cortex. To preview, while perceptual expectations were reflected in the deep layers[38], high confidence false percepts (i.e., hallucinations) were in contrast reflected in the middle layers of V2, and were not driven by perceptual expectations. The presence of these high confidence false percepts correlated with hallucination severity in a larger sample. In other words, perceptual expectations and hallucinations elicited orthogonal laminar responses in the early visual cortex, suggesting that hallucinations can arise from spontaneous stimulus-specific activity in sensory input layers.

Results

Experimental procedure

In the layer-specific fMRI study, 25 healthy participants were asked to complete a difficult perceptual discrimination task, in which they were required to report the orientation (45° or 135°) of a grating that was embedded in visual noise, as well as report their confidence that a grating was present (Fig. 1a). Unbeknownst to the participants, an auditory cue predicted the orientation of the upcoming gratings (Fig. 1b). On trials in which a grating was presented (grating-present trials, 50%), the auditory cues were 100% valid. Crucially, on 50% of trials the gratings were omitted, and only visual noise was presented. Our research questions pertained to participants' behaviour on these grating-absent trials. Specifically, we investigated whether participants reported high confidence false percepts on these trials, and if so, whether these were reflected in the feedback or input layers of the early visual cortex.

Participants experienced false percepts that were independent of perceptual expectation cues

Participants' accurately identified the grating orientation on grating-present trials more often than expected by chance (mean=0.83, SD=.09) (T{24}=18.1, *p*<.001). Furthermore, they were more accurate when they were confident that they had seen a grating (i.e., higher than average confidence across trials) than when they were not (high: mean=0.90, SD=0.09; low: mean=0.75, SD=.12; paired t-test) (T{24}=5.7, *p*<.001) (Fig. 1c), demonstrating that they were able to perform the task and used the confidence ratings in a meaningful way. Participants also reported the perceived orientation more quickly on grating-present than grating-absent trials (F{1,24}=12.10, *p*=.002), as well as when they were more confident that they had seen a grating (F{1,24}=21.04, *p*<.001). Participants were more confident on grating-present trials (mean=2.49, SD=.60) than grating-absent trials (mean=2.19, SD=0.54) (T{24}=4.76, *p*<.001) (Fig. 1d). Upon debriefing, all participants but one underestimated the frequency of grating-absent trials, believing on average that .14 (SD=.13) of trials contained just noise, while the true proportion was .50 (Fig. 1e). Strikingly, participants reported perceiving a grating with high confidence (3 out of 4 or higher) on 36% of grating-absent trials (Fig. 1f). Surprisingly, the perceptual expectation cues did not significantly bias which orientation participants perceived on grating-absent trials (0.53 false percepts congruent with the cue, chance level is 0.50, T{24}=1.90, *p* = 0.07). The small numerical trend towards false percepts being congruent with the expectation cues was driven by a few individuals who became aware of the meaning of the cues (N = 7 out of 25; Fig. 1g), potentially reflecting a concomitant response bias. High confidence, i.e. hallucinated, percepts were not more affected by the cues than low confidence, i.e. guessed, percepts (T{23}= 0.37, *p*=.71). Trial by trial predictors of participants' choice behaviour were explored using a logistic regression model (see Online Methods for details). As expected, orientation responses on grating-present trials were predominantly driven by the presented stimulus (T{21}=12.6, *p*<.001), but also by which orientation was perceived on the previous trial (T{21}=5.3, *p*<.001). Interestingly, on grating-absent trials previous percepts also significantly predicted orientation reports (T{23}=5.48, *p*<.001), whereas the cues did not (T{23}=2.03, *p*=.056) (Fig. 1h). As above, the trend towards the cues influencing perception on grating-absent trials was driven by a few participants who became aware of the meaning of the cues. Thus, previous percepts were a more important contributor to false percepts than the expectation cues were.

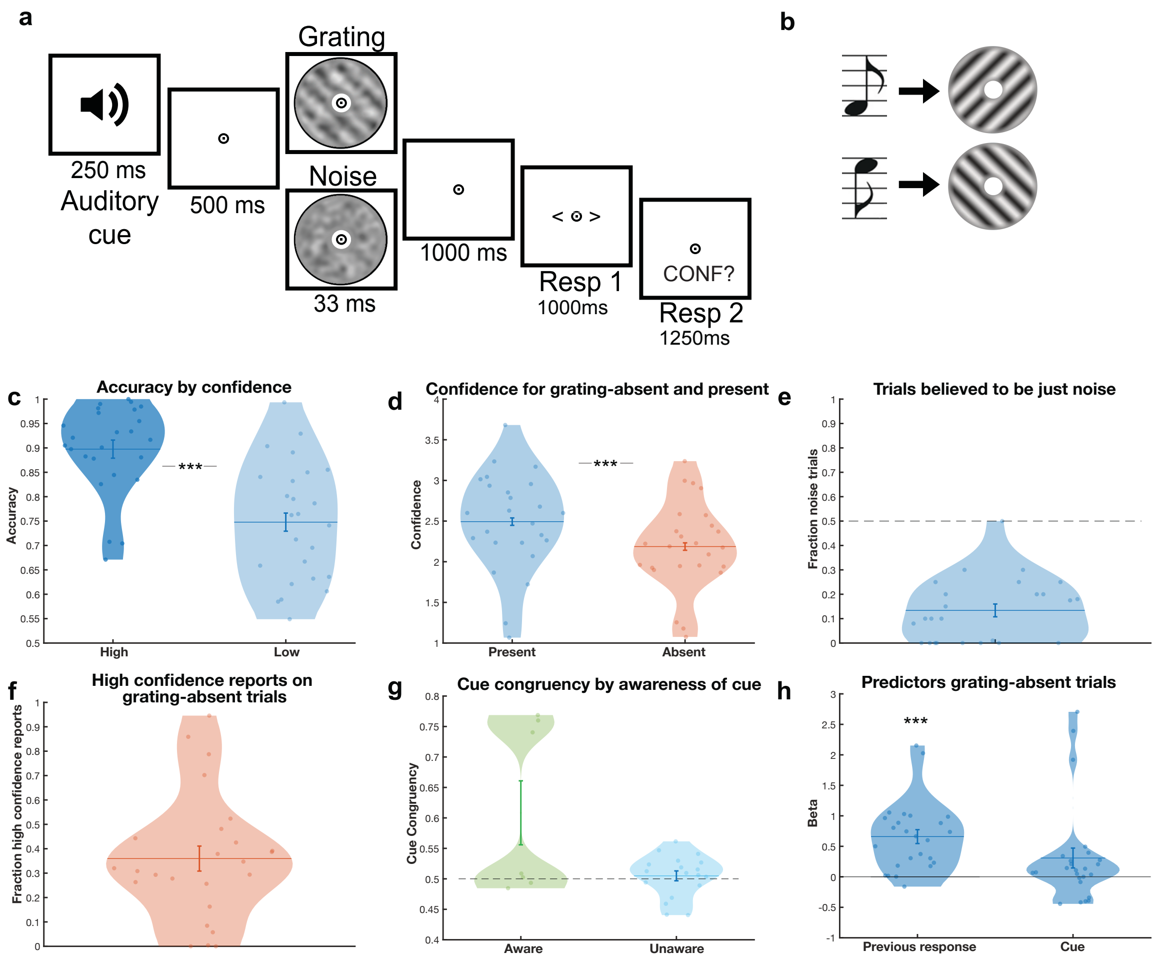

Fig. 1 | Experimental design and behavioural findings of layer-specific fMRI study. a*, During the experiment an auditory cue was followed by either a low contrast grating embedded in noise (50% of trials), or a noise patch (50%). Participants indicated which orientation they saw and how confident they were that a grating was presented.* ***b****, One sound predicted the appearance of a 45 degree, or clockwise, oriented grating, whilst the other predicted a 135 degree, or anti-clockwise, orientated grating. Auditory cues were 100% valid on grating-present trials.* ***c****, On grating-present trials, participants' orientation responses were more accurate when they indicated that they were confident (dark blue) compared to not confident (light blue) that they had seen a grating.* ***d****, Participants were more confident on grating-present trials (blue) compared to grating-absent trials (orange).* ***e****, Participants on average believed only ~14% of trials to contain just noise (whereas the true proportion was 50%).* ***f****, On grating-absent trials, participants reported seeing gratings with high confidence (3 out of 4 or higher) on an average of 36% of trials.* ***g****, There was a slightly higher, non-significant, tendency to report orientations congruent with the expectation cue on grating-absent trials, which was driven by a few participants who were aware of the cue.* ***h****, Participants' orientation response on the previous trial significantly predicted their orientation response on grating-absent trials*** p<.001. Dots represent individual participants and curved shapes indicate density. Error bars indicate within subject standard error of the mean for figure c & d, and standard error of the mean for e, f, g & h.*

Estimation of layer-specific activity

Ultra high-field (7T) fMRI was used to estimate stimulus-specific activity in the deep, middle, and superficial layers of the early visual cortex (V1 and V2) [38–40]. To examine orientation-specific blood oxygen level–dependent (BOLD) activity, V1 and V2 voxels were divided into two (45°-preferring and 135°-preferring) subpopulations based on an independent functional localiser, and layer-specific BOLD profiles were estimated for these subpopulations separately. Orientation-specific laminar BOLD profiles were calculated by subtracting the laminar profile obtained from orientation-incongruent voxels (e.g., 135°-preferring voxels when a 45° grating was perceived) from the laminar profile in orientation-congruent voxels (45°-preferring voxels in this example). Orientation-specific BOLD profiles induced by false percepts and expectations were estimated on trials where the gratings were omitted, and only visual noise was presented (see Online Methods for details).

Stimulus-specific activity reflecting false percepts and expectations in distinct cortical layers

Our main hypotheses pertained to the layer-specific activity induced by false percepts and perceptual expectations. In V2, perceptual expectations, false percepts with high confidence (interpreted as hallucinations), and those with low confidence (interpreted as guesses) generated stimulus-specific activity in different cortical layers (F{4,96}=3.82, *p*=.006; Figure 2). High and low confidence false percepts were subsequently compared directly, revealing a significant difference in laminar profiles (F{2,48}=4.56, *p*=.015). This was driven by increased stimulus-specific activity for high compared to low confidence false percepts in the middle layers (T{24}=2.80, p=.010), but not in the superficial (T{24}=0.68, *p*=.50) or deep (T{24}=-0.12, *p*=.92) layers (Fig. 2a). In short, only high confidence false percepts were significantly reflected in stimulus-specific activity in the middle layers (T{24}=3.19, *p*=.002, corrected: *p*=.006). A direct comparison of high confidence false percepts and perceptual expectations demonstrated that they were associated with different laminar profiles (F{2,48}=5.218, *p*=.009). This was primarily driven by a difference in middle layer activity (T{24}=3.58, *p*=.002), which were activated by high confidence false percepts (T{24}=3.19, *p*=.002, corrected: *p*=.006) but not by expectations (T{24}=-0.54, *p*=.70). Furthermore, perceptual expectations significantly activated deep layers (Fig. 2a) (T{24}=2.46, *p*=.011, corrected: *p*=.033, Wilcoxon: *p*=.005, corrected *p*=.014), whereas hallucinations did not (T{24}=0.48, *p*=.32), although the difference was not significant (T{24}=0.56, *p*=.59). Superficial layers were not significantly activated by perceptual expectations (T{24}=0.58, *p*=.28) or hallucinations (T{24}=1.20, *p*=.12), and there was no significant difference in superficial layer activity in hallucinations and perceptual expectations (T{24}=0.75, *p*=.46). In sum, high confidence false percepts were reflected by stimulus-specific activity in the middle layers, whereas perceptual expectations selectively activated the deep layers. Interestingly, the effect of perceptual expectations in the deep layers was not driven by those participants who became aware of the meaning of the cues (T{23}=0.32, *p*=.78), and was significantly present in the subset of participants who were not aware of the cue meanings (N=18, T{17}=1.87, *p*=.039, see supplementary figure 2). This suggests that the brain can generate sensory expectations in the early visual cortex based on implicit associations, without necessarily generating conscious experiences associated with them.

No effects were found in V1 (all *p*>.1). In addition, there was an interaction between layer (superficial, middle, and deep), stimulus condition (high and low confidence false percepts, and perceptual expectations), and ROI (V1 and V2) (F{4,96}=3.33, *p*=.013). This demonstrates that the layer-specific activity for the different stimulus conditions was different for the two ROIs, which in combination with the null findings in V1 shows that the effects were specific to V2. This is likely explained by the relatively low spatial frequency (0.5 cycles/°) gratings used here being more effective in activating V2 than V1[41]. In line with this, a cross-validated analysis of orientation-specific BOLD signals within the functional localiser revealed stronger orientation-specific effects in V2 than V1 (F{1,23}=10.36, *p*=.004; Supplementary Fig. 2a). Note that the low-contrast gratings embedded in noise on the 50% grating-present trials did not evoke significant orientation-specific BOLD activity in either V1 or V2 (all *p*>.1). It is perhaps not surprising that such weak and noisy stimuli did not evoke a significant orientation-specific BOLD signal, but it is striking to note that more cognitive processes like expectation and perception do.

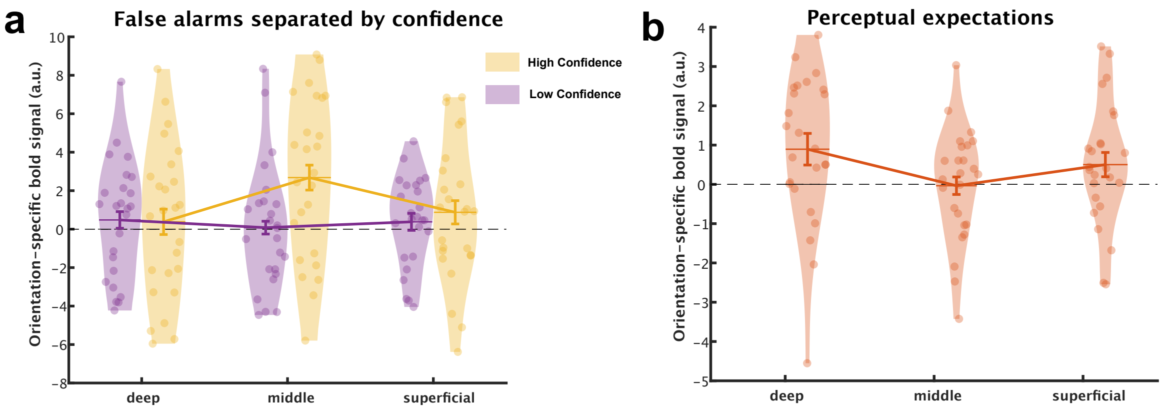

***Fig. 2 | Orientation-specific BOLD effects in the cortical layers of V2. a**, The middle layers contained orientation-specific activity reflecting high confidence false percepts, whilst low confidence trials did not induce orientation-specific activity in any of the layers. **b**, The deep layers reflected orientation-specific activity induced by perceptual expectations on grating-absent trials (red).*

High confidence false percepts and reduced sensory precision predict everyday hallucination severity

In a separate online experiment (N=100), we tested whether the high confidence false percepts that were related to middle layer activity in the layer-specific fMRI study correlated with the prevalence and severity of hallucinatory percepts in daily life, as measured by the Cardiff Anomalous Perception Scale (CAPS) questionnaire[42]. The false percept task used here was virtually identical to the one used in the fMRI experiment (with slight variations in practice procedure and trial counts; see Online Methods for details). One important difference was the introduction of three (rather than one) contrast levels on the grating-present trials, to enable estimates of sensory precision, i.e., how task accuracy depended on evidence quality. Specifically, a base-level contrast value was selected for each participant based on their performance during the practice phase, and this base-contrast was used during the main experiment along with gratings with 1% higher and 1% lower contrast.

Crucially, the prevalence of abnormal perceptual experiences in daily life (total CAPS scores[42], was positively correlated with the average confidence that participants reported on grating-absent trials – i.e., the prevalence of hallucinated gratings in our task – across participants in the online sample (Rho=.22, *p*=.029 (Fig. 3a). Further, the sensory precision term – the influence of grating contrast on choice behaviour – correlated negatively with abnormal perceptual experience scores (Rho=-.30, *p*=.003) (Fig. 3b). In other words, the less sensitive participants were to stimulus contrast, the more likely they were to experience abnormal perceptual experiences in daily life. The effect of the expectation cues was not correlated with abnormal perceptual experience severity (*p*>.1). In general, the expectation cues had a very minor effect on choice behaviour, as in the fMRI experiment (see Supplementary Results). Using a linear regression model, both confidence on grating-absent trials (T{99}=2.01, *p*=.048) and sensory precision (T{99}=-2.98, *p*=.004) were found to be separate predictors of abnormal perceptual experience severity (overall linear regression model: F{2,97}=6.12, *p*=.003, R^2=0.112). We did not find a relation between average confidence on grating-absent trials and delusion ideation (Rho=.13, *p*=.20), but there was a correlation with the sensory precision term (Rho=-.29, *p*=.004). In sum, high confidence false percepts, which were reflected in stimulus-specific activity in the middle layers of V2 in the fMRI study, were related to the severity of everyday abnormal perceptual experiences. This suggests that spontaneous stimulus-like activity in the middle layers may contribute to the experience of abnormal perception in everyday life.

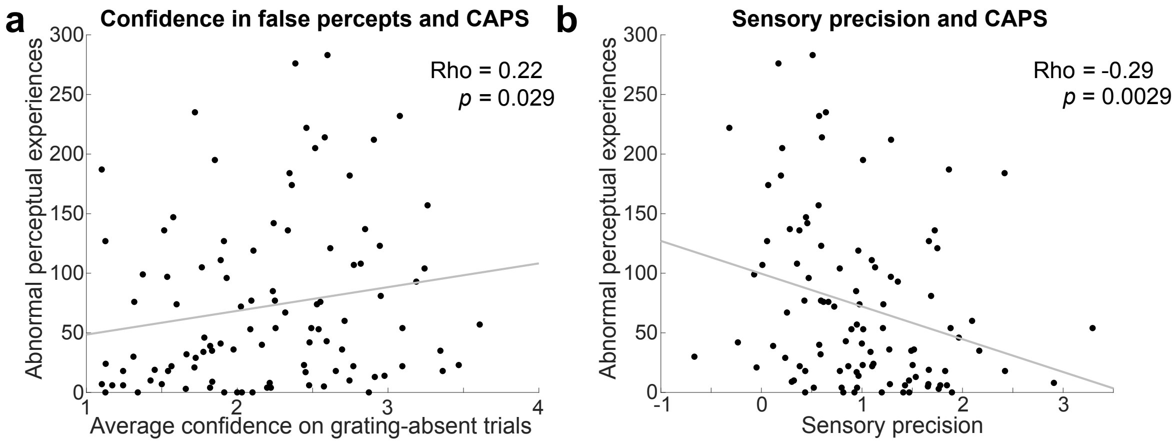

Fig. 3 | Correlations between grating percepts in the current task and everyday hallucinations. a*, Abnormal perceptual experiences correlated with confidence in false percepts.* ***b****, Abnormal perceptual experiences correlated with less reliance on sensory precision (i.e., the interactive effect of grating contrast and orientation on choice behaviour).*

Discussion

There is a wide range of theories that attempt to explain the neural mechanisms of hallucinations. The dominant theory proposes that hallucinations are driven by excessive feedback activity conveying perceptual expectations[1–3], whereas others have suggested a crucial role for feedforward activity[17]. Here, these two hypotheses were tested against each other using layer-specific fMRI. We found that the deep layers of the visual cortex reflected perceptual expectations, replicating previous research[38]. However, hallucinations, defined as high confidence false percepts, were reflected in the middle input layers of the early visual cortex, and were unaffected by perceptual expectation cues. These high confidence false percepts correlated with everyday hallucination severity in a separate online study. This suggests that hallucinations can arise from stimulus-specific feedforward like activity in the sensory cortex, and do not require stimulus-specific feedback activity in the deep layers of the visual cortex.

The core finding here – that orientation-specific activity in the middle layers of V2 can lead participants to perceive a grating that was not actually presented – has important implications for the field of hallucination research as well as perception research more generally. That is, it suggests that hallucinations can arise through activity in the input layers, in the absence of top-down effects of stimulus-specific perceptual expectations. This puts a larger emphasis on feedforward signals than previously assumed in dominant models of hallucinations, which have largely emphasised the role of top-down expectations[1–3,7,11,12].

These findings are in line with theories that state that hallucinations can arise from spontaneous activity that resembles sensory input[17], which recent studies have confirmed in neurological disorders like Charles Bonnet Syndrome[16]. Furthermore, they expand upon previous studies that have found that early sensory activity can lead to false alarms[26–30,43], by suggesting that the activity reported by these studies may reflect spontaneous fluctuations in the input layers, rather than feedback from higher-order regions. Recent circular inference models of hallucinations have emphasised the role of ascending loops, akin to feedforward activity, in unimodal hallucinations as seen in psychotic disorders[20,44]. Specifically, they suggest that weak sensory signals can trigger perceptual hypotheses that are then counted as sensory evidence themselves in runaway 'overcounting' loops. This overcounting of sensory signals has been shown to correlate with positive symptoms (e.g., hallucinations and delusions) in schizophrenia patients[21].

Top-down processes like visual working memory and perceptual expectations induce stimulus representations in the deep and superficial, but not the middle layers of the early visual cortex[38,40,45]. However, keeping an image in visual working memory or merely expecting a visual stimulus does not lead to a concurrent perceptual experience, as is the case with a hallucination. It might therefore not be surprising that hallucinations are not linked to agranular layer activity, but instead require the middle layers to be active. That is, activity in the middle layers may be necessary to generate a perceptual experience that is attributed externally. Indeed, it has been suggested that feedforward signals are essential in distinguishing imagination from veridical perception[46].

The prevalence of high confidence false percepts, i.e., hallucinations as defined in our experiment, correlated with everyday abnormal perceptual experiences. This is in line with previous studies demonstrating that those who hallucinate are more prone to perceive stimuli in noise in detection tasks[11,12,14,47]. Furthermore, higher hallucination scores correlated negatively with the sensory precision term of our logistic regression model. That is, the less participants relied on the contrast of the sensory stimulus in making their perceptual decision, the more they experienced hallucinations in everyday life. This is in line with Bayesian models of hallucinations, where a reduction in sensory precision increases the influence of prior expectations, possibly leading to hallucinations[48]. Interestingly, reduced reliance on sensory contrast on the one hand, and confidence on grating-absent trials on the other, were separate predictors of everyday hallucinations severity, suggesting separate underlying mechanisms contributing to hallucinations. The present study revealed that activity fluctuations in the middle input layers reflected hallucinations. The strength of such fluctuations in an individual, in combination with a loss of precision in sensory encoding in these same input layers, might particularly amplify the tendency to experience hallucinations in everyday life. This is in line with the idea that a loss of sensory precision is a separate compounding factor, which allows hallucinations to manifest[49] . In psychiatric conditions, this might come in the form of internal noise in the sensory systems. In neurological disorders, this is most prominently seen in Charles Bonnet syndrome[17,50].

These findings should not be taken as evidence against the theory that top-down perceptual expectations can play an important part in generating hallucinations, for which there is ample indirect evidence[10–15,47,51]. In fact, perceptual expectations might still play a role in the present study through the experimental set-up. Specifically, on every trial there was a strong expectation that one of two stimuli would be presented. This is underlined by the debriefing questionnaire, where participants greatly underestimated the proportion of trials on which the gratings were absent, suggesting that they expected to see gratings on most trials, regardless of their specific orientation. This expectation about stimulus presence versus absence could be an important driver of hallucinations. Indeed, it is typically the expectation of the likelihood of the presence of a stimulus, rather than its specific contents, that has been found to be associated with psychosis in previous studies[11–13]. Interestingly, expectations about stimulus presence versus absence and expectations about stimulus content have been suggested to be supported by different neural processes[30,52,53].

On a neural level, an expectation of stimulus presence might prime pyramidal neurons in the deep cortical layers through receptors on their apical dendrites[54,55]. Targeting the apical dendrites is expected not to drive these pyramidal neurons directly, but allow them to function as coincidence detectors[55]. In turn, deep layer neurons can modulate incoming sensory input through their projections onto the middle layers[56,57]. This allows sensory input concurrent with expectations to be processed more quickly, giving them a head start in signal processing[55,58,59]. Furthermore, once evidence accumulates for one orientation in the middle layers, possibly due to random fluctuations, the evidence for the other orientation may be suppressed through lateral inhibition[60,61]. This hypothesised circuit has the potential to implement circular inference, whereby top-down expectations can get counted as sensory evidence in ascending loops[20,21,62]. This would subsequently result in stimulus-specific activity in the middle layers driving hallucinations as seen in our study. This modulatory (rather than driving) role of feedback connections[57] may explain why many studies have found that concurrent noisy input is required to induce hallucinations[9–14,47]. Alternatively, the prior on stimulus presence may not have affected sensory cortex itself, but instead biased source monitoring processes in the prefrontal cortex[46] towards judging spontaneous fluctuations in the middle layers of V2 as being externally generated.

An alternative possibility is that there are two different types of hallucinations: feedforward driven hallucinations, and hallucinations driven by top-down perceptual expectations. These might reflect phenomenologically different hallucinations, potentially mapping onto so-called minor phenomena[63] and complex visual hallucinations[64] respectively. There is some evidence that these reflect different neural processes in Lewy body dementia[65]. The hallucinations reported in the current study may be more akin to minor phenomena, reflecting low-level stimulus-specific activity resulting in low-level visual distortions. Those driven by perceptual expectations could result in more complex visual hallucinations, and might in contrast be reflected by deep layer activity. Future studies should aim to study the laminar profile of hallucinations elicited by perceptual expectations to test these hypotheses.

Why did the expectation cues not induce hallucinations in the current study? We suggest this may be due to recruiting a normative sample who might rely less on expectations than those prone to hallucinate[9,11,14]. Second, there is increasing evidence that conscious expectations exert stronger effects on perception than unconscious expectations[66,67]. Since for the majority of individuals in this experiment the expectations were implicit, this might have weakened their effect. Studies that do report effects of implicit cues on perception typically reveal biased perception of existing stimuli, rather than eliciting percepts de novo[68,69] although see[70] for a notable exception.

Despite the expectation cues not affecting perception, they did induce stimulus-specific templates in the deep layers of the early visual cortex, in line with previous work[38]. Strikingly, this effect was reliable even in those who were not aware of the cue-stimulus relationship, which suggests that the brain can generate sensory expectations based on statistical relationships that are learnt outside of conscious awareness[71] (Aitken & Kok, 2022). The finding that stimulus-specific deep layer activity might not be sufficient to generate a conscious percept seems in conflict with theories that emphasise the importance of deep layer activity in generating conscious percepts[72,73]. Our work nuances that picture, by suggesting that deep layer activity by itself may not be sufficient for conscious awareness. Of course, we cannot rule out that the deep layer activity revealed here was simply not strong enough to induce conscious percepts.

In the present study, no reliable effects of either perceptual expectations or false percepts were found in V1. This is likely due to the lower spatial frequency stimuli used in the present study (0.5 cycles/°) compared to previous studies that reported stimulus-specific effects in V1 (1.0-1.5 cycles/°;[39,40,74]), as V2 neurons prefer lower spatial frequencies than V1 neurons[75]. This was confirmed by our localiser analyses, showing stronger stimulus-specific effects in V2 than in V1. While the localiser task did induce stimulus-specific effects in V2, no stimulus-driven effects were found for the presented stimuli in the main task, likely due to these being very low contrast and embedded in noise. This suggests that neural signals induced by expectations and hallucinated gratings are more reliable than those evoked by very noisy bottom-up inputs, in line with a previous study that also found stimulus-specific activity reflecting false percepts, but not presented stimuli embedded in noise[29].

In conclusion, hallucinated percepts were reflected by stimulus-specific activity in the middle input layers of the early visual cortex, whereas perceptual expectations activated the deep layers. These findings suggest that hallucinations can arise from low-level content-specific fluctuations in the input layers of the visual cortex. This nuances the view that hallucinations are necessarily driven by top-down expectations[1–3]. Future studies should aim to further explore the nature of these low-level fluctuations and what drives them, as well as investigate whether hallucinations can also be driven by purely top-down signals. These findings shed light on the neural mechanisms underlying hallucinations, revealing how the brain can generate perception in the absence of sensory input.

Acknowledgements

The authors would like to thank Fraser Aitken for help with early piloting of the experimental paradigm. This work was supported by a Wellcome/Royal Society Sir Henry Dale Fellowship [218535/Z/19/Z] and a European Research Council (ERC) Starting Grant [948548] to P.K. The Wellcome Centre for Human Neuroimaging is supported by core funding from the Wellcome Trust [203147/Z/16/Z].

Competing interests

The authors declare no competing interests

Online Methods

Ethics statement

This study was approved by the University College London Research Ethics Committee (R13061/RE002 for the imaging study, R6649/RE004 for the online study) and was conducted according to the principles of the Declaration of Helsinki. All Participants gave written informed consent prior to participation and received monetary compensation (£7.50 an hour for behavioural tasks, £10 an hour for MRI).

Participants

Twenty-eight healthy human volunteers with normal or corrected-to-normal vision participated in the 7T fMRI experiment. Three participants were excluded due to our strict head motion criteria of no more than 10 movements larger than 1.0 mm in any direction between successive functional volumes. For the remaining participants, the maximum change in head position in any direction over the course of the fMRI runs was within 4 mm (0.66 +/− 0.54 mm, mean +/− SD over participants) of the mean head position (to which the anatomical boundaries were registered). The final sample consisted of 25 participants (22 female; age 25 ± 4 years; mean ± SD).

One hundred participants participated in the online study. Participants were recruited through Prolific (www.prolific.co) and were paid £7.50 for their participation. Three participants were excluded for failing to answer the catch questions on the questionnaire correctly, resulting in a final sample of 97 participants.

Questionnaires

For the online study questionnaire data was collected for the Peter Delusions Index[76], as well as the Cardiff Abnormal Perception Scale[42]. Total scores were calculated for the PDI and CAPS by adding their respective subscales. These were then correlated with the behavioural measures for the online study.

Stimuli

Grayscale luminance-defined sinusoidal Gabor grating stimuli were generated using MATLAB (MathWorks, Natick, Massachusetts, United States of America, RRID:SCR_001622) and the Psychophysics Toolbox (Brainard, 1997). During the behavioural session for the fMRI study, the stimuli were presented on a PC (1920 × 1200 screen resolution, 60-Hz refresh rate). In the fMRI scanning session, stimuli were projected onto a rear projection screen using an Epson EB-L1100U Laser projector (1920 × 1200 screen resolution, 60-Hz refresh rate) and viewed via a mirror (viewing distance 91 cm). On grating-present trials (50%), auditory cues were followed by a grating (0.5-cpd spatial frequency, 33-ms duration, and separated by a 750-ms blank screen), displayed in an annulus (outer diameter: 10° of visual angle, inner diameter: 1°, contrast decreasing linearly to 0 over 0.7° at the inner and outer edges), surrounding a fixation bull's eye (0.7° diameter). These stimuli were combined with one of 4 noise patches, which resulted in a 4% contrast grating embedded in 20% contrast noise during the fMRI session. On grating-absent trials, one of the 4 noise patches was presented on its own. Noise patches were created through smoothing pixel-by-pixel Gaussian noise with a Gaussian smoothing filter, ensuring that the spatial frequency of the noise patches matched that of the gratings. This was done to ensure that the noise patches and gratings had similar low-level properties, increasing the likelihood of reporting false percepts. To avoid including noise patches which contained grating-like orientation signals by chance, the noise patches were processed through a number of Gabor energy filters with varying preferred orientations. The noise patches with low (2%) signal energy were selected to be included for the present experiment. The resulting 4 noise patches were used for all participants throughout the experiment, ensuring that reported false percepts could only be triggered by internal mechanisms[29,77]. Importantly, the 4 noise patches did not elicit consistent biases in orientation responses (all *p*>.05), confirming that the method was successful in generating noise patches that did not resemble either orientation. During the practice session on the first day, the contrast of the gratings was initially high (80%), gradually decreasing to 4% towards the end of the practice. The central fixation bull's-eye was present throughout the trial, as well as during the intertrial interval (ITI; jittered exponentially between 2,150 and 5,150 ms). In the online study, multiple grating contrast levels were presented, titrated to each individual (see procedure below).

Experimental procedure

On the first day of testing for the laminar fMRI study, participants underwent a behavioural practice session. The practice consisted of an instruction phase with 7 blocks of 16 trials where the task was made progressively more difficult, whilst verbal and written instructions were provided. During the practice runs, the auditory cues predicted the orientation of the first grating stimulus of the pair with 100% validity (45° or 135°; no grating-absent trials). After the completion of the instructions, the participants completed 4 runs of 128 trials each, separated into 2 blocks of 64 trials each. In the first 2 runs the expectation cues were 100% valid, to ensure participants learnt the association, whilst in the final 2 runs the cue was 75% valid, to test whether participants might have adopted a response bias. Grating contrast decreased over the 4 runs, specifically the contrast levels were 7.5, 6, 5, and 4%, while the contrast of the noise patches remained constant at 20%. No grating-absent trials were presented on day 1. On the second day, participants performed the same task in the MR scanner. As on the first day, 4 runs were completed, but now the grating contrast was fixed at 4% on grating-present trials, and on 50% of the trials the gratings were omitted and only noise patches were presented, resulting in grating-absent trials. Each run lasted ~12.5 minutes, totalling ~50 minutes.

Trials consisted of an auditory expectation cue, followed by a grating stimulus embedded in noise on 50% of trials (750-ms stimulus onset asynchrony (SOA) between cue and grating). The auditory cue (high or a low tone) predicted the orientation of the grating stimulus (45° or 135°). On grating-present trials, a grating with the orientation predicted by the auditory cue was presented embedded in noise, while on grating-absent trials only a noise patch was presented. The stimulus was presented for 33ms in the fMRI study. After the stimulus disappeared, the orientation response prompt appeared, consisting of a left and right pointing arrow on either side of the fixation dot (location was counterbalanced). Participants were required to select the arrow corresponding to their answer (left arrow for anti-clockwise, or 135°, right arrow for clockwise, or 45°; 1s response window) through a button press with their right hand. Subsequently the letters "CONF?" appeared on the screen probing participants to indicate their confidence that they had seen a grating (1 = I did not see a grating, 2 = I may have seen a grating, 3 = I probably saw a grating, 4 = I am sure I saw a grating), using 1 of 4 buttons with their left hand (1.25s response window). Participants indicated their response using an MR-compatible button box in the MRI scanner, and a keyboard during training.

After the main experiment, participants performed a functional localiser task inside the scanner. This consisted of flickering gratings (2 Hz), presented at 100% contrast, in blocks of approximately 14.3 seconds (4 TRs). Each block contained gratings with a fixed orientation (45° or 135°). The 2 orientations were presented in a pseudorandom order followed by an approximately 14.3-second blank screen, containing only a fixation bull's-eye. Participants were tasked with responding whenever the black fixation dot briefly dimmed to ensure central fixation. All participants were presented with 16 localiser blocks which totalled approximately 15 minutes.

The online study was created and hosted using Gorilla Experiment Builder (www.gorilla.sc)[78]. Participants were recruited through Prolific (www.prolific.co). Prior to the start of the instruction blocks participants were asked to keep a 50cm distance from the screen. They were required to adjust the size of a rectangle on their screen to match a bank card, to ensure that the visual angle was equal across participants. Subsequently, they were asked to adjust the volume of their headphones to a high but not unpleasant volume. The instruction phase of the online study was the same as the instruction phase of the fMRI study. The timings for the trials were identical as well, except that the stimuli appeared on the screen for 50ms instead of 33ms, due to software constraints. After completing the 7 instruction blocks, participants were required to complete 4 blocks with different grating contrast levels (7.5, 6, 5, and 4%, in that order). The lowest grating contrast for which participants were able to perform the orientation task with at least 75% accuracy served as the base contrast value for the main experiment. During the main experiment, participants were required to complete 4 blocks of 128 trials. Unlike in the fMRI experiment, different grating contrast levels were presented, and expectation cues were sometimes invalid (6.7% of grating present trials). Specifically, out of the 128 trials on each block, on 96 trials (75%) a grating was present and on 32 (25%) only a noise patch was presented. Of the 96 grating-present trials, one third (32) were presented at the base contrast, one third at base - 1% contrast, and the other third at base + 1% contrast. Of these grating-present trials, 90 (93.3 %) were valid and 6 (6.7 %) were invalid.

fMRI data acquisition

MRI data were acquired on a Siemens Magnetom Terra 7T MRI system (Siemens Healthcare GmbH, Erlangen, Germany) with an 8 channel head coil for localised transmission, operating in a quadrature-like ('TrueForm') mode, with a 32-channel head coil insert for reception (Nova Medical, Wilmington, USA) at the Wellcome Centre for Human Neuroimaging (University College London). Functional images were acquired using a T2*-weighted 3D gradient-echo EPI sequence (volume acquisition time of 3,552 ms, TR = 74 ms, TE = 26.95 ms, voxel size 0.8 × 0.8 × 0.8 mm^3, 15° flip angle, field of view 192 × 192 × 38.4 mm^3, GeneRalized Autocalibrating Partial Parallel Acquisition (GRAPPA) acceleration factor 4, and partial Fourier 6/8 in the phase-encoded direction of the EPI readout, binomial (1331) water-selective excitation). Anatomical images were acquired using a Magnetization Prepared 2 Rapid Acquisition Gradient Echo (MP2RAGE) sequence (TR = 5,000 ms, TE = 2.54 ms, TI = 900 ms and 2,750 ms, voxel size 0.65 × 0.65 × 0.65 mm, 5° and 3° flip angles, field of view 208 × 208 × 156 mm^3, in-plane GRAPPA acceleration factor 3).

Preprocessing of fMRI data

The first 2 volumes of each run were discarded. Prior to registration the functional volumes were cropped to cover only the occipital lobe to reduce the influence of severe distortions in the frontal lobe. The cropped functional volumes were spatially realigned within scanner runs, and subsequently between runs, to correct for head movement, using SPM12 (https://www.fil.ion.ucl.ac.uk/spm/).

Segmentation and coregistration of cortical surfaces

The methods for segmenting and coregistering cortical surfaces are identical to several previously published studies[34,38–40], and are reiterated here. Freesurfer (http://surfer.nmr.mgh.harvard.edu/) was used to detect the grey (GM) and white matter (WM) boundaries and cerebral spinal fluid (CSF) based on a bias corrected MP2RAGE image. The boundaries were checked for errors where the dura was mistakenly included in the pial surface. Subsequently the GM boundaries were registered to the mean functional image. Specifically, a conventional rigid-body registration was followed by a recursive boundary-based registration (RBR)[79]. RBR consisted of applying boundary-based registration (BBR) recursively to increasingly smaller partitions of the cortical mesh. An affine BBR was applied with 7 degrees of freedom: rotation and translation along all 3 dimensions and scaling along the phase-encoding direction only. This scaling allows correction of distortions along the low bandwidth phase-encoded EPI direction of acquisition. In each iteration, the cortical mesh was split into 2, and the optimal BBR transformations were found and applied to the respective parts. Subsequently, each part was split into 2 again and registered. The specificity increased at each stage and corrected for local mismatches between the structural and the functional volumes that are due to magnetic field inhomogeneity-related distortions. Six such iterations were performed. The splits were made along the cardinal axes of the volume, such that the number of vertices was equal for both parts. The plane for the second cut was orthogonal to the first, the third was orthogonal to the first 2. The median displacement was taken after running the recursive algorithm 6 times, in which different splitting orders where used, comprised of all 6 permutations of x, y, and z.

Definition of regions of interest

The definition of regions of interests (ROIS) was identical to a recently published study[38], and reiterated here. V1 and V2 surface labels were obtained through Freesurfer, based on the segmentation of the MP2RAGE image. These were subsequently projected to volume space, covering the full cortical depth plus a 50% extension into WM and CSF. The V1 and V2 ROIs were subsequently constrained to the voxels that were responsive to the localiser gratings. Specifically, separate regressors were defined for the blocks of 45° and 135° gratings, respectively, and the mean of the resulting parameter estimates was contrasted against baseline to identify voxels that exhibited a significant response to the grating stimuli irrespective of orientation ($T > 2.3$, $p < 0.05$; V1: mean=6208, SD=1799 voxels; V2: mean=9370, SD=3256 over participants). Subsequently, the orientation preference of each voxel was estimated by contrasting the 2 orientation regressors. The 500 voxels that most strongly favoured the 45° and 135° gratings, respectively, constituted the two orientation-specific ROIs within V1 and V2. Finally, each voxel's time course was normalised (z-scored), and multiplied by the absolute T-value of the orientation contrast (45° versus 135°), to weight the data by the most robust orientation preference. Note that all reported effects in the z-scored data were also present without z-scoring. These ROI definitions were identical to those used in previous studies that successfully resolved orientation-specific BOLD signals with layer specificity [38–40]. The analysis approach was matched to these previous studies to facilitate comparisons between previous findings that involve orientation- and layer-specific fMRI signals.

Definition of the cortical layers

GM was divided into 3 equivolume layers using the level set method described in detail elsewhere[79–81], following the principle that the layers of the cortex maintain their volume ratio throughout the curves of the gyri and sulci [81]. Briefly, the level set function is a signed distance function (SDF), where points on the same surface equal 0 and values on one side of the surface are negative and values on the other are positive. The level set function for the GM–CSF and GM–WM boundaries is calculated, and then intermediate surfaces can be defined by moving the surface to intermediate cortical depths. The equivolume model transforms a desired volume fraction into a distance fraction, taking the local curvature of the pial and WM surfaces at each voxel into account [79]. Two intermediate surfaces between the WM and pial boundaries were calculated, yielding 3 GM layers (deep, middle, and superficial). In human early visual cortex, these 3 laminar compartments are expected to correspond roughly to layers I to III, layer IV, and layers V and VI, respectively [82]. Based on these surfaces, 4 SDFs were calculated, containing for each functional voxel its distance to the boundaries between the 5 compartments (WM, CSF, and the 3 GM layers). This set of SDFs (or "level set") allowed the calculation of the distribution of each voxel's volume over the 5 compartments [79]. This layer volume distribution provided the basis for the laminar GLM discussed below.

Extraction of layer-specific time courses

Because the fMRI data consisted of 0.80 mm isotropic voxels, individual voxels will naturally contain signals from multiple layers, as well as WM and CSF. Thus, if we were to simply interpolate the fMRI signal at different depths, there will be contamination from bordering layers. One way to address this so-called partial volume problem is to decompose the layers by means of a spatial GLM [34,35,38,40,79]. For every ROI, a laminar design matrix **X** represents the distribution of the 500 voxels over the different layers (n x k, where n = 500 voxels, and k = 5 laminar compartments). Every row of **X** indicates the proportions of the layers covered by a particular voxel, and the columns represent the volume of the corresponding layer across voxels. This laminar design matrix can be used in a spatial GLM to separate the BOLD signal of the 5 different laminar compartments (three GM layers, WM, and CSF) through ordinary least squares (OLS) regression [79]:

$$Y = X \cdot B + \varepsilon$$

Here **Y** is a vector of voxel values from an ROI in a specific functional volume, **X** is the laminar design matrix, and **B** is a vector of layer signals. For each ROI and each functional volume, the layer signal $\widehat{\boldsymbol{B}}$ was estimated by regressing **Y** against **X**, yielding 5 depth-specific time courses per ROI.

To confirm that the method correctly identified GM, the raw signal in the EPI volumes for each of the 3 GM layers was quantified, as well as WM and CSF. As expected, the signal intensity was higher in the 3 GM layers (deep: 239 +/− 38; middle: 240 +/− 46; superficial: 239 +/− 40; mean +/− SD over participants) than in WM (209 +/− 30) and CSF (230 +/− 51) (T{24}= 5.69, p = 7.4 × 10^{-6}).

Estimating effects of interest per layer

A temporal GLM was used to estimate the effects of interest in each of the 3 GM layers. A model with 4 regressors of interest was used to estimate the effects of perceptual expectation (grating-present and grating-absent trials, separately for the 45° or 135° orientations). These regressors of interest were constructed by convolving stick functions representing the onsets of the trials with SPM12's canonical haemodynamic response function as well as their temporal derivative, resulting in Beta values for each experimental effect. Furthermore, the head motion parameters, their derivatives, and the square of the derivatives, were included as nuisance regressors. Subsequently, the data and the design matrix were high-pass filtered (cut-off = 128 seconds) to remove any low-frequency signal drifts.

In order to calculate orientation-specific BOLD responses, the layer-specific parameter estimates for each orientation in the non-corresponding ROI (e.g., a 45° grating/expectation in a 135°-preferring ROI) were subtracting from the parameter estimates in their corresponding ROI (e.g., a 45° grating/expectation in a 45°-preferring ROI; see equation below where B stands for Beta).

$$OrientationSpecificEffect = (45B45ROI + 135B135ROI) - (45B135ROI + 135B45ROI)$$

This procedure was followed for all the laminar analyses presented in this study (perceptual expectations, high confidence false percepts, low confidence false percepts). These estimated BOLD responses were subjected to a 2-way repeated measures ANOVA with factors perceptual condition (expectation, high confidence false percept, low confidence false percept), and cortical layer (deep, middle, superficial). The main effect of interest, namely whether laminar BOLD profiles differed for perceptual expectations and hallucinated gratings, was tested by the interaction of perceptual condition and cortical layer. To follow up a significant effect with all 3 perceptual conditions included, further repeated measures effects were performed to specifically test 1) the interaction between perceptual expectation vs high confidence false percept and cortical layer to explore whether hallucinations were specifically different from perceptual expectations, and 2) the interaction between high and low confidence false percepts and cortical layer to explore whether being confident in a false percept affects the laminar profile. Significant interactions were followed up with paired-sample *t* tests. Finally, orientation-specific effects in specific layers were tested against zero using one-sample t-tests (one-tailed). To visualise the relevant across-subject variance for the within-subject ANOVA, errors bars in all figures show within-subject standard error of the mean (SEM)[83,84].

Behavioural analyses

For the online study, accuracy and confidence scores were compared across the different contrast levels using repeated measures ANOVAs. Accuracy was also compared between the different confidence levels, to test whether participants were more accurate at identifying grating orientation when they were more confident that they had seen a grating. The effect of the expectation cues was assessed by exploring whether participants tended to report orientations in line with the cue. Follow-up tests were performed to investigate whether the cues' effects were mediated by awareness of their meaning. To understand what drives abnormal perceptual experiences, a logistic regression model was used to explore which factors predicted orientation responses on grating-present and grating-absent trials separately. Predictors for grating-present trials were current stimulus orientation, current stimulus contrast, orientation predicted by the cue, orientation response on the previous trial, and the interaction between present stimulus contrast and orientation (as a measure of sensory precision). For the grating-absent trials, the predictors included previous orientation response, and orientation predicted by the cue (as there was no present stimulus orientation or contrast). Finally, we tested whether abnormal perceptual experiences as measured using

the CAPS questionnaire were correlated with cue effects, confidence on grating-absent trials, and sensory precision, using Spearman's rank correlation.
Similarly, for the fMRI study we probed the modulation of accuracy by confidence, the proportion of high confidence false percepts on grating-absent trials, and the proportion of cue congruent responses. Participants' orientation responses were also explored with a logistical regression model, but without stimulus contrast as a predictor as this was not varied for the purposes of the fMRI experiment.

1. Powers, A. R., Kelley, M. & Corlett, P. R. Hallucinations as Top-Down Effects on Perception. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging* vol. 1 393–400 (2016).
2. Sterzer, P. *et al.* The Predictive Coding Account of Psychosis. *Biological Psychiatry* vol. 84 634–643 (2018).
3. Corlett, P. R. *et al.* Hallucinations and Strong Priors. *Trends in Cognitive Sciences* vol. 23 114–127 (2019).
4. Reichert, D. P., Seriès, P. & Storkey, A. J. Charles Bonnet Syndrome: Evidence for a Generative Model in the Cortex? *PLoS Computational Biology* **9**, (2013).
5. Friston, K. The free-energy principle: a rough guide to the brain? *Trends in Cognitive Sciences* **13**, 293–301 (2009).
6. Bastos, A. M. *et al.* Canonical Microcircuits for Predictive Coding. *Neuron* vol. 76 695–711 (2012).
7. Haarsma, J., Kok, P. & Browning, M. The promise of layer-specific neuroimaging for testing predictive coding theories of psychosis. *Schizophrenia Research* (2020) doi:10.1016/j.schres.2020.10.009.
8. de Lange, F. P., Heilbron, M. & Kok, P. How Do Expectations Shape Perception? *Trends in Cognitive Sciences* vol. 22 764–779 (2018).
9. Schmack, K. *et al.* Delusions and the role of beliefs in perceptual inference. *Journal of Neuroscience* **33**, 13701–13712 (2013).
10. Teufel, C. *et al.* Shift toward prior knowledge confers a perceptual advantage in early psychosis and psychosis-prone healthy individuals. *Proc Natl Acad Sci U S A* **112**, 13401–13406 (2015).
11. Powers, A. R., Mathys, C. & Corlett, P. R. *Pavlovian conditioning-induced hallucinations result from overweighting of perceptual priors*. https://www.science.org.
12. Kafadar, E. *et al.* Modeling perception and behavior in individuals at clinical high risk for psychosis: Support for the predictive processing framework. *Schizophrenia Research* **226**, 167–175 (2020).
13. Schmack, K., Bosc, M., Ott, T., Sturgill, J. F. & Kepecs, A. Striatal dopamine mediates hallucination-like perception in mice. *Science (1979)* **372**, (2021).
14. Haarsma, J. *et al.* Influence of prior beliefs on perception in early psychosis: Effects of illness stage and hierarchical level of belief. *Journal of Abnormal Psychology* **129**, 581–598 (2020).
15. Cassidy, C. M. *et al.* A Perceptual Inference Mechanism for Hallucinations Linked to Striatal Dopamine. *Current Biology* **28**, 503-514.e4 (2018).
16. Hahamy, A., Wilf, M., Rosin, B., Behrmann, M. & Malach, R. How do the blind "see"? The role of spontaneous brain activity in self-generated perception. *Brain* **144**, 340–353 (2021).

17. Burke, W. The neural basis of Charles Bonnet hallucinations: A hypothesis. *Journal of Neurology Neurosurgery and Psychiatry* **73**, 535–541 (2002).

18. Desai, N. S., Rutherford, L. C. & Turrigiano, G. G. *Plasticity in the intrinsic excitability of cortical pyramidal neurons*. http://neurosci.nature.com (1999).

19. Painter, D. R., Dwyer, M. F., Kamke, M. R. & Mattingley, J. B. Stimulus-Driven Cortical Hyperexcitability in Individuals with Charles Bonnet Hallucinations. *Current Biology* **28**, 3475-3480.e3 (2018).

20. Denève, S. & Jardri, R. Circular inference: Mistaken belief, misplaced trust. *Current Opinion in Behavioral Sciences* vol. 11 40–48 (2016).

21. Jardri, R., Duverne, S., Litvinova, A. S. & Denève, S. Experimental evidence for circular inference in schizophrenia. *Nature Communications* **8**, (2017).

22. Schmack, K., Schnack, A., Priller, J. & Sterzer, P. Perceptual instability in schizophrenia: Probing predictive coding accounts of delusions with ambiguous stimuli. *Schizophrenia Research: Cognition* **2**, 72–77 (2015).

23. Teufel, C., Kingdon, A., Ingram, J. N., Wolpert, D. M. & Fletcher, P. C. Deficits in sensory prediction are related to delusional ideation in healthy individuals. *Neuropsychologia* **48**, 4169–4172 (2010).

24. Notredame, C. E., Pins, D., Deneve, S. & Jardri, R. What visual illusions teach us about schizophrenia. *Frontiers in Integrative Neuroscience* vol. 8 (2014).

25. Weilnhammer, V. *et al.* Psychotic experiences in schizophrenia and sensitivity to sensory evidence. *Schizophrenia Bulletin* **46**, 927–936 (2020).

26. Boly, M. *et al. Baseline brain activity fluctuations predict somatosensory perception in humans*. www.pnas.orgcgidoi10.1073pnas.0611404104 (2007).

27. Busch, N. A., Dubois, J. & VanRullen, R. The phase of ongoing EEG oscillations predicts visual perception. *Journal of Neuroscience* **29**, 7869–7876 (2009).

28. Wyart, V. & Tallon-Baudry, C. How ongoing fluctuations in human visual cortex predict perceptual awareness: Baseline shift versus decision bias. *Journal of Neuroscience* **29**, 8715–8725 (2009).

29. Pajani, A., Kok, P., Kouider, S. & de Lange, F. P. Spontaneous activity patterns in primary visual cortex predispose to visual hallucinations. *Journal of Neuroscience* **35**, 12947–12953 (2015).

30. Podvalny, E., Flounders, M. W., King, L. E., Holroyd, T. & He, B. J. A dual role of prestimulus spontaneous neural activity in visual object recognition. *Nature Communications* **10**, (2019).

31. Felleman, D. J. & van Essen, D. C. *Distributed Hierarchical Processing in the Primate Cerebral Cortex*. https://academic.oup.com/cercor/article/1/1/1/408896.

32. Harris, K. D. & Mrsic-Flogel, T. D. Cortical connectivity and sensory coding. *Nature* vol. 503 51–58 (2013).

33. Muckli, L. *et al.* Contextual Feedback to Superficial Layers of V1. *Current Biology* **25**, 2690–2695 (2015).

34. Kok, P., Bains, L. J., van Mourik, T., Norris, D. G. & de Lange, F. P. Selective activation of the deep layers of the human primary visual cortex by top-down feedback. *Current Biology* **26**, 371–376 (2016).

35. Lawrence, S. J. D., Formisano, E., Muckli, L. & de Lange, F. P. Laminar fMRI: Applications for cognitive neuroscience. *NeuroImage* vol. 197 785–791 (2019).

36. Stephan, K. E. *et al.* Laminar fMRI and computational theories of brain function. *NeuroImage* vol. 197 699–706 (2019).

37. Self, M. W., van Kerkoerle, T., Goebel, R. & Roelfsema, P. R. Benchmarking laminar fMRI: Neuronal spiking and synaptic activity during top-down and bottom-up processing in the different layers of cortex. *NeuroImage* vol. 197 806–817 (2019).

38. Aitken, F. *et al.* Prior expectations evoke stimulus-specific activity in the deep layers of the primary visual cortex. *PLoS Biology* **18**, (2020).

39. Lawrence, S. J., Norris, D. G. & de Lange, F. P. Dissociable laminar profiles of concurrent bottom-up and top-down modulation in the human visual cortex. (2019) doi:10.7554/eLife.44422.001.

40. Lawrence, S. J. D. *et al.* Laminar Organization of Working Memory Signals in Human Visual Cortex. *Current Biology* **28**, 3435-3440.e4 (2018).

41. Foster, K. H., Gaska, J. P., Naglert, M., Pollen, D. A. & Foster And, K. H. *SPATIAL AND TEMPORAL FREQUENCY SELECTIVITY OF NEURONES IN VISUAL CORTICAL AREAS Vi AND V2 OF THE MACAQUE MONKEY*. *J. Physiol* vol. 365 (1985).

42. Bell, V., Halligan, P. W. & Ellis, H. D. The Cardiff Anomalous Perceptions Scale (CAPS): A new validated measure of anomalous perceptual experience. *Schizophrenia Bulletin* **32**, 366–377 (2006).

43. Ress, D. & Heeger, D. J. Neuronal correlates of perception in early visual cortex. *Nature Neuroscience* **6**, 414–420 (2003).

44. Leptourgos, P., Bouttier, V., Denève, S. & Jardri, R. From hallucinations to synaesthesia: A circular inference account of unimodal and multimodal erroneous percepts in clinical and drug-induced psychosis. *Neuroscience & Biobehavioral Reviews* **135**, 104593 (2022).

45. van Kerkoerle, T., Self, M. W. & Roelfsema, P. R. Layer-specificity in the effects of attention and working memory on activity in primary visual cortex. *Nature Communications* **8**, (2017).

46. Dijkstra, N., Kok, P. & Fleming, S. M. Perceptual reality monitoring: Neural mechanisms dissociating imagination from reality. *Neuroscience & Biobehavioral Reviews* 104557 (2022) doi:10.1016/j.neubiorev.2022.104557.

47. Stuke, H., Kress, E., Weilnhammer, V. A., Sterzer, P. & Schmack, K. Overly Strong Priors for Socially Meaningful Visual Signals Are Linked to Psychosis Proneness in Healthy Individuals. *Frontiers in Psychology* **12**, (2021).

48. Adams, R. A., Stephan, K. E., Brown, H. R., Frith, C. D. & Friston, K. J. The computational anatomy of psychosis. *Frontiers in Psychiatry* vol. 4 (2013).

49. Landgraf, S. & Osterheider, M. "To see or not to see: That is the question." The "Protection-Against-Schizophrenia" (PaSZ) model: Evidence from congenital blindness and visuo-cognitive aberrations. *Frontiers in Psychology* vol. 4 (2013).

50. ffytche, D. *et al.* The anatomy of conscious vision: an fMRI study of visual hallucinations. *Nature Neuroscience* **1**, 738–742 (1998).

51. Zarkali, A. *et al.* Increased weighting on prior knowledge in Lewy body-associated visual hallucinations. *Brain Communications* **1**, (2019).

52. Mazor, M., Friston, K. J. & Fleming, S. M. Distinct neural contributions to metacognition for detecting, but not discriminating visual stimuli. *Elife* **9**, (2020).

53. Samaha, J., Iemi, L., Haegens, S. & Busch, N. A. Spontaneous Brain Oscillations and Perceptual Decision-Making. *Trends in Cognitive Sciences* vol. 24 639–653 (2020).

54. Spruston, N. Pyramidal neurons: Dendritic structure and synaptic integration. *Nature Reviews Neuroscience* vol. 9 206–221 (2008).

55. Larkum, M. A cellular mechanism for cortical associations: An organizing principle for the cerebral cortex. *Trends in Neurosciences* vol. 36 141–151 (2013).
56. Binzegger, T., Douglas, R. J. & Martin, K. A. C. A quantitative map of the circuit of cat primary visual cortex. *Journal of Neuroscience* **24**, 8441–8453 (2004).
57. Kim, J., Matney, C. J., Blankenship, A., Hestrin, S. & Brown, S. P. Layer 6 corticothalamic neurons activate a cortical output layer, layer 5a. *Journal of Neuroscience* **34**, 9656–9664 (2014).
58. Antic, S. D., Zhou, W. L., Moore, A. R., Short, S. M. & Ikonomu, K. D. The decade of the dendritic NMDA spike. *Journal of Neuroscience Research* vol. 88 2991–3001 (2010).
59. Major, G., Larkum, M. E. & Schiller, J. Active properties of neocortical pyramidal neuron dendrites. *Annual Review of Neuroscience* vol. 36 1–24 (2013).
60. Hawkins, J. & Ahmad, S. Why Neurons Have Thousands of Synapses, a Theory of Sequence Memory in Neocortex. *Frontiers in Neural Circuits* **10**, (2016).
61. Hawkins, J., Ahmad, S. & Cui, Y. A theory of how columns in the neocortex enable learning the structure of the world. *Frontiers in Neural Circuits* **11**, (2017).
62. Jardri, R. *et al.* Are hallucinations due to an imbalance between excitatory and inhibitory influences on the brain? *Schizophrenia Bulletin* **42**, 1124–1134 (2016).
63. Pagonabarraga, J. *et al.* Minor hallucinations occur in drug-naive Parkinson's disease patients, even from the premotor phase. *Movement Disorders* **31**, 45–52 (2016).
64. Walterfang, M., Velakoulis, D. & Mocellin, R. *Neuropsychiatry of complex visual hallucinations*. *Australian and New Zealand Journal of Psychiatry* vol. 40 (2006).
65. D'Antonio, F. *et al.* Visual hallucinations in Lewy body disease: pathophysiological insights from phenomenology. *Journal of Neurology* (2022) doi:10.1007/s00415-022-10983-6.
66. Meijs, E. L., Slagter, H. A., de Lange, F. P. & van Gaal, S. Dynamic interactions between top–down expectations and conscious awareness. *Journal of Neuroscience* **38**, 2318–2327 (2018).
67. Alilović, J., Slagter, H. A. & van Gaal, S. Subjective visibility report is facilitated by conscious predictions only. *Consciousness and Cognition* **87**, (2021).
68. Aitken, F., Turner, G. & Kok, P. Prior Expectations of Motion Direction Modulate Early Sensory Processing. *The Journal of Neuroscience* **40**, 6389–6397 (2020).
69. Kok, P., Brouwer, G. J., van Gerven, M. A. J. & de Lange, F. P. Prior expectations bias sensory representations in visual cortex. *Journal of Neuroscience* **33**, 16275–16284 (2013).
70. Chalk, M., Seitz, A. R. & Seriès, P. Rapidly learned stimulus expectations alter perception of motion. *Journal of Vision* **10**, (2010).
71. Aitken, F. & Kok, P. Hippocampal representations switch from errors to predictions during acquisition of predictive associations. *Bioarchives* 1–31 (2022).
72. Aru, J., Suzuki, M., Rutiku, R., Larkum, M. E. & Bachmann, T. Coupling the State and Contents of Consciousness. *Frontiers in Systems Neuroscience* **13**, (2019).
73. Takahashi, N. *et al.* Active dendritic currents gate descending cortical outputs in perception. *Nature Neuroscience* **23**, 1277–1285 (2020).
74. Kok, P., Failing, M. F. & de Lange, F. P. Prior expectations evoke stimulus templates in the primary visual cortex. *Journal of Cognitive Neuroscience* **26**, 1546–1554 (2014).
75. Foster, K. H., Gaska, J. P., Naglert, M., Pollen, D. A. & Foster And, K. H. *SPATIAL AND TEMPORAL FREQUENCY SELECTIVITY OF NEURONES IN VISUAL CORTICAL AREAS Vi AND V2 OF THE MACAQUE MONKEY*. *J. Physiol* vol. 365 (1985).

76. Peters, E., Joseph, S., Day, S. & Qarety, P. *Measuring Delusional Ideation: The 21-Item Peters et aL Delusions Inventory (PDI)*. https://academic.oup.com/schizophreniabulletin/article/30/4/1005/1930847.
77. Wyart, V., Nobre, A. C. & Summerfield, C. Dissociable prior influences of signal probability and relevance on visual contrast sensitivity. *Proc Natl Acad Sci U S A* **109**, 3593–3598 (2012).
78. Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N. & Evershed, J. K. Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods* **52**, 388–407 (2020).
79. van Mourik, T., van der Eerden, J. P. J. M., Bazin, P. L. & Norris, D. G. Laminar signal extraction over extended cortical areas by means of a spatial GLM. *PLoS ONE* **14**, (2019).
80. Kleinnijenhuis, M. *et al.* Diffusion tensor characteristics of gyrencephaly using high resolution diffusion MRI in vivo at 7T. *Neuroimage* **109**, 378–387 (2015).
81. Waehnert, M. D. *et al.* Anatomically motivated modeling of cortical laminae. *Neuroimage* **93**, 210–220 (2014).
82. de Sousa, A. A. *et al.* Comparative cytoarchitectural analyses of striate and extrastriate areas in hominoids. *Cerebral Cortex* **20**, 966–981 (2010).
83. Cousineau, D. Confidence intervals in within-subject designs: A simpler solution to Loftus and Masson's method. *Tutorials in Quantitative Methods for Psychology* **1**, 42–45 (2005).
84. Morey, R. D. *Confidence Intervals from Normalized Data: A correction to Cousineau (2005)*. *Tutorial in Quantitative Methods for Psychology* vol. 4 (2008).