1 Title Page

Title: Dynamic predictive templates in perception

Authors:

Veith Weilnhammer[1], Yuki Murai[2], David Whitney[1]

Affiliations and Contribution:

[1] Helen Wills Neuroscience Institute, University of California Berkeley, USA

[2] Center for Information and Neural Networks, Osaka, Japan

Corresponding Author:

Veith Weilnhammer, Helen Wills Neuroscience Institute, University of California Berkeley, USA, email: veith.weilnhammer@gmail.com

Word Count: 2242 words (main text without methods)

2 Summary

False alarms occur when participants report a signal that was not presented. Hallucinations reflect perceptual experiences that unfold in the absence of an external cause. To serve as an experimental proxy for hallucinations, false alarms must meet two criteria: First, false alarms should not only reflect erroneous reports of a signal, but represent perceptual states that are characterized by specific contents. Second, false alarms should occur on a timescale compatible with the temporal dynamics of hallucinations. In this work, we combined a classification image approach with Hidden Markov Models to show that false alarms can match the perceptual and temporal characteristics of hallucinations. We asked healthy human participants to discriminate difficult-to-detect visual stimuli from noise, and found that false alarms were more likely to occur during an internal mode of sensory processing, a minute-long state of the brain during which the content of perception is strongly biased toward previous experiences. Our results suggest that hallucinations are driven by dynamic predictive templates that transform noise into transient, recurring, coherent, and meaningful perceptual experiences.

3 Results

Hallucinations are vivid and transient experiences of objects, such as images or sounds, that occur in the absence of a corresponding stimulus[1–7]. These phenomena occur across a wide spectrum of neuropsychiatric disorders, but are particularly characteristic of schizophrenia, a severe and chronic mental health condition typically associated with repeated episodes of psychosis[6,8,9].

To better understand the computational and neural mechanisms of hallucinations, cognitive neuroscience has primarily focused on experiments that induce false alarms (FA) in healthy participants[1–5,7], psychosis prone individuals[1,3,4], and patients diagnosed with schizophrenia[5].

FAs occur when participants make decisions about difficult-to-detect stimuli, and report the presence of a signal that was, in fact, not presented.

Previous studies have shown that FA become more likely when people *expect* the presence of a stimulus. In line with this idea, FAs can be induced via cross-modal conditioning[1,3–5] and manipulations of signal probability[7]. A number of empirical findings indicate that such conditioned FAs may provide a window into the neural mechanisms of hallucinations: Conditioned FAs are more frequent in hallucination-prone individuals[1–4,7] and patients diagnosed with schizophrenia[5]. Moreover, they can be triggered by dopamine[7], an established molecular endophenotype of schizophrenia[10], as well as by ketamine[7], a NMDA-receptor antagonist that elicits psychotic symptoms[11].

However, one of the most important limitations in the current state-of-the art is that FAs are, at heart, reports. What remains to be established is whether FAs are accompanied by perceptual experiences. Even though hallucinations have varying phenomenological properties across the disorders in which they occur, and may reflect unstructured noise-like percepts, particularly at early stages of psychotic illness, they are typically associated with specific perceptual contents that have meaning to the afflicted individual[12–14]. Moreover, the focus on behavioral reports has led previous studies to investigate FAs at the level of individual stimulus presentations within signal detection experiments[1,7]. This assumes that the underlying mechanisms are isolated events that depend, at least in part, on an experimental intervention[1,3–5,7], trait[1–5,7] or diagnosis[5]. Hallucinations, by contrast, are neither momentary nor constant, but typically unfold as coherent and transient experiences over the course of seconds, minutes or hours[15,16].

Here, we present a novel paradigm that can help to determine whether FAs share the perceptual and temporal characteristics of hallucinations. First, we use a classification image approach[17] to show that FAs cannot be reduced to behavioral reports, but occur when the external sensory input matches the features of an internal predictive template that is defined

3

by a specific perceptual content (criterion 1). Second, we apply Hidden Markov Models[18] (HMMs) to show that FA, and the underlying predictive templates, are linked to recurring minute-long intervals of internal mode processing, a state of the brain during which the content of perception is strongly biased toward previous experiences (criterion 2).

3.1 Criterion 1: The perceptual quality of FAs

According to Bayes theorem, alarms ($P(signal|input)$) become likely when people expect to see a signal ($P(signal)$, prior), and when the features of a noisy input are compatible with the expected signal ($P(input|signal)$, likelihood). In natural environments, where the recent past is a predictor of the near future, previous stimuli may induce expectations about what is likely to be perceived next[19,20]. This generates predictive templates that integrate experiences over time[21] ($P(signal|input) = P(signal) \times P(input|signal)$), giving rise to the phenomenon of perceptual serial dependence[22–26].

Predictive templates may improve the interpretation of noisy but predictable sensory inputs[21,27]. When overly strong, however, predictive templates lose their adaptive function and may cause people to experience spurious signals in noise ($P(signal|noise)$). With respect to the study of hallucinations, the critical question is whether FAs are driven solely by signal expectations ($P(signal)$, prior), in which case they may just be behavioral responses without any perceptual quality, or if they also depend on the compatibility of the noise with the expected signal ($P(noise|signal)$, likelihood). The latter case would indicate that FAs are accompanied by perceptual experiences with a specific content.

To test the contribution of prior and likelihood to FAs, we analyzed data from 22 participants who judged whether close-to vertical gratings (signals) were present (alarms) or absent (rejections) in white noise image (Figure 1A). The alarm rate depended on three factors: the contrast of the signal at the current trial (4.38 ± 0.07, z = 60.76, p $< 2.2 \times 10^{-308}$), the contrast of the signal at the preceding trial (0.11 ± 0.03, z = 4.33, p $= 1.5 \times 10^{-5}$, Figure

76 2), and the relative power of the current noise image at close-to-vertical orientations (60

77 to 120 degrees), relative to the total power over all orientations (0.1 ± 0.01, z = 8.97, p =

78 2.9×10^{-19}, Figure 1B).

79 This confirms that FAs were driven by predictive templates: FAs were more likely when the

80 stimulus presented at the preceding trial induced a strong signal expectation ($P(signal)$),

81 and when the sensory noise at the current trial matched the features of the expected signal

82 ($P(noise|signal)$). To visualize these predictive templates, we subtracted the noise power at

83 rejection trials from the noise power at alarm trials (Figure 3). The resulting classification

84 images[17] revealed a noise power peak at close-to-vertical-orientations (0.1 ± 0.02, $\mathrm{T}(1.37 \times 10^4)$

85 = 7.24, p = 4.84×10^{-13}) that matched the average orientation of the signals.

86 If FAs were simply lapses or biases in response behavior, one would expect a flat classification

87 image that mirrors the overall flat power-by-orientation distribution in the white noise stimuli

88 used in our experiment. This was clearly not the case: FAs occurred more frequently at noise

89 stimuli that matched a feature of the signal (high power at close-to-vertical orientations;

90 Figure 3). Critically, observers learned the features of the average signal from the sequence

91 of experiences they made throughout the experiment.

92 Our findings therefore support criterion 1: at the time of FAs, observers perceived specific

93 contents - in this case, gratings with a close-to-vertical orientation. The hallucinated content

94 was determined by predictive templates that are generated by the sequence of preceding

95 experiences[21–26] ($P(signal|noise) = P(signal) \times P(noise|signal)$).

3.2 Criterion 2: The temporal dynamics of predictive templates

97 Hallucinations are transient, coherent, an often-meaningful experiences that unfold over

98 time[12–16]. A valid experimental proxy for hallucinations should therefore fluctuate at a

99 timescale compatible with the duration of hallucinatory experiences. In the case of FAs, this

100 means that the underlying predictive templates should be dynamic, and generate recurrent

and transient intervals during which perception is strongly biased toward previous experiences.

To assess the temporal dynamics of the predictive templates, we estimated HMMs that inferred transitions between two latent states, each linked to a General Linear Model (GLM) that predicted trial-wise responses y_t (alarms versus rejections) from the signal contrast at the current trial ($\beta_s \times s_t$) and the response at the preceding trial ($\beta_H \times y_{t-1}$, Figure 1C). In line with previous results[18,20], the HMMs revealed slow alternations between an external mode, where perception closely followed external stimulus ($\beta_S > \beta_H$), and an internal mode, where perception was strongly biased by preceding experiences ($\beta_S < \beta_H$, Figure 1C).

Bayesian model comparison indicated that the two-state GLM provided a more parsimonious description of the data than the one-state control model (reduction in BIC: $\delta_{BIC} = -1.12 \times 10^3$; see Supplemental Figure S1 for control analyses on randomly permuted data, for which the advantage of the two-state GLM was lost).

Participants spent $28.7 \pm 3.5\%$ of their time in internal mode, with alternation between modes occurring in intervals of 59.1 ± 8.8 trials, corresponding to 118.21 ± 17.6 sec. External and internal mode fluctuated spontaneously, and were only marginally modulated by contrast of the signal grating (Supplemental Figure S2). The GLMs that defined the external and internal mode were independent from the orientation of the signal, the contrast of the signal at the preceding trial, and the power-by-orientation distribution of the noise. Any between-mode difference in the classification images was therefore indicative of dynamic changes in the underlying predictive templates.

We found that the internal mode increased the rate of FAs relative to external mode (0.1 ± 0.02, T(109) $= 4.61$, p $= 1.1 \times 10^{-5}$, Figure 2). In the internal mode, perception depended more on the contrast of the stimulus presented at the preceding trial (1 ± 0.06, z $= 16.86$, p $= 8.76 \times 10^{-64}$) and less on contrast of the stimulus presented at the current trial (-1.16 ± 0.15, z $= -7.64$, p $= 2.18 \times 10^{-14}$). The effect of signal expectations ($P(signal)$) on perception were thus exaggerated during the internal mode.

Alarms during the internal mode were associated with a close-to-vertical noise-power-peak that was shifted toward the orientation of the stimulus presented at the preceding trial (19.87 ± 6.73, T(20.28) = 2.95, p = 7.82×10^{-3}, Figure 3). This shift was not observed during the external mode (−3.67 ± 6.27, T(20.15) = −0.59, p = 0.56). This suggests that the predictive templates, and the associated bias in the content of perception toward previous experiences, were particularly strong during the internal mode.

Our findings therefore confirmed criterion 2: the predictive templates that drive FAs are dynamic and fluctuate according to two opposing external and internal modes of perception. Alternations between modes occurred at the order of minutes, and were thus compatible with the temporal duration of hallucinations. During the internal mode, FAs were associated with hallucinated contents that were skewed toward previous experiences.

4 Discussion

Our results indicate that the FAs observed here are dynamic perceptual phenomena, which are characterized by specific content and are driven by dynamic predictive templates. These templates fluctuate in intervals of several minutes and are particularly strong during internal modes of sensory processing. This confirms that FAs induced by signal-detection experiments can match the perceptual[12–14] and temporal[15,16] quality of hallucinations.

Our findings suggest that FAs, and potentially hallucinations, are linked to alternations between external and internal modes of perception. External and internal modes have repeatedly been observed in mice[20,28], healthy human participants[20], and patients diagnosed with schizophrenia[18]. We found that the internal mode not only increased the rate of FAs, but also enhanced the extent to which the content perceived at the time of FAs was skewed toward previous experiences.

This finding points to a tentative mechanism for how hallucinations emerge, persist, and end: During the internal mode, a hallucination may emerge as a FA that occurs when (i), a specific

context induces signal expectations (e.g., hearing the neighbors voice in one's apartment, $P(signal)$) and (ii), the sensory input is compatible with the average features of the expected signal (e.g., construction noise that matches the features of human speech, $P(input|signal)$). The internal mode enhances the effect of preceding experiences on perception and may allow the hallucinated content to persist over time as a coherent experience. A transition from internal to external mode greatly reduces the strength of the predictive template in which the hallucination unfolds, ultimately causing the end of the experience. Future research could test this hypothesis by testing whether FAs, and ultimately the hallucinations experienced by people living with psychotic symptoms, can be mitigated by interventions that disrupt the internal mode.

During the internal mode, the average vectors of the classification images were displaced by 20.18 ± 3.12 degrees from vertical (Figure 3). The direction of the vectors depended on the predictive template, i.e., the orientation of the stimulus presented at the preceding trial. Interestingly, the displacement of the classification vectors from vertical was larger than the orientation of the high-contrast gratings (-10 and 10 degrees from vertical, Figure 1A), which were most effective at inducing FAs during the internal mode (Figure 2). Since the high-contrast gratings stood out from the other stimuli in terms of contrast and orientation (low contrast gratings were presented at orientation ranging from -3 to 3, Figure 1A-B), one may suspect that the perceived orientation of the high-contrast grating exceeded their veridical orientation. An alternative explanation may be that FAs experiences exaggerated the features of the inducer, just as hallucinations often portray exaggerated version of everyday stimuli (such as hallucinated speech that stands out in terms of tone, pitch, or extreme content[12–14]).

Our findings suggests a key role of serial dependence[22–25] and predictive templates[26] for how hallucinations emerge and persist over time. This view overlaps with the so-called strong prior account of schizophrenia[9], where psychotic symptoms, and in particular hallucinations,

178 are thought to occur due to an exaggerated effect of prior predictions on perception[1–5,7].

179 At the same time, they seem at odds with the observation that prior are less effective in

180 psychosis prone individuals[29] and patients diagnosed with schizophrenia[30] (the weak prior

181 account of schizophrenia[8]). So far, attempts to reconcile these disparate set of findings have

182 argued that priors may be either weak or strong depending on the stage of psychotic illness[9]

183 (weak priors in early stages, and strong priors in later stages) or the cognitive hierarchy[9]

184 (weak priors at the perceptual level, and strong priors at the cognitive level).

185 The fact that FAs, and potentially psychotic symptoms such as hallucinations, are tied

186 to recurrent alternation between external and internal modes of perception may provide

187 an alternative explanation for the apparent discrepancy between strong and weak priors:

188 Psychosis prone individuals and patients diagnosed with schizophrenia may share general

189 propensity toward the external mode, which reduces the effects of priors on perception[29]

190 and drives erratic responses to ambiguous sensory information[18]. During the internal mode,

191 psychosis prone individuals and patients diagnosed with schizophrenia may have stronger

192 predictive templates that trigger and shape the content of hallucinations. The combination of

193 classification images and HMMs provides a novel analytical framework to test this hypothesis,

194 may help to advance our theoretical understanding of schizophrenia, and can pave the way

195 towards innovations in mental health that benefit people living with psychotic symptoms.

5 Methods

5.1 Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Veith Weilnhammer (veith.weilnhammer@gmail.com).

5.2 Materials availability

This study did not generate new unique reagents.

5.3 Data and code availability

All custom code and behavioral data are available on https://github.com/veithweilnhammer/ Modes. This manuscript was created using the *R Markdown* framework, which integrates all data-related computations and the formatted text within one document. With this, we wish to make our approach fully transparent and reproducible for reviewers and future readers.

5.4 Experimental model and subject details

5.4.0.1 Apparatus The participants viewed visual stimuli on a gamma-corrected CRT monitor (Sony Trinitron Multiscan G520, 1024x768 pixels, 100 Hz refresh rate, Konica Minolta LS-110 luminometer for gamma correction) at a viewing distance of 57.3 cm, with head stabilization provided by a chin rest.

5.4.1 Stimuli and Procedure

All stimuli were presented using MATLAB (MathWorks, R2017a) and the Psychophysics Toolbox. In 44% of the trials, we presented low-contrast near-to-vertical Gabor patches which were embedded in static white noise. In another 44% of the trials, we presented only static white noise. The remaining 12% of the trials featured a high-contrast Gabor patch (40%

10

Michelson contrast) embedded in white noise. This high-contrast inducer was introduce to boost the phenomenon of serial dependency. The inducers were oriented 10 degrees clockwise or counterclockwise relative to vertical, and presented in intervals randomized between 4 and 10 trials. Stimuli were displayed for 500 ms, with a maximum white noise contrast of 60%. To reduce potential luminance aftereffects, the signal was masked by low-pass filtered luminance contrast noise (100% contrast) for 500 ms, and the Gabor phase was randomized across trials.

Visual stimuli measured 14 x 14 degrees of visual angle (d.v.a.), with Gabor patches confined to a Gaussian contrast envelope with a 3 d.v.a. standard deviation. The spatial frequency of the Gabor patches was 0.5 cycles per degree (cpd), with orientation randomized within ± 3 degrees relative to vertical for low-contrast gratings. High-contrast gratings were presented at ± 10 degrees. Participants were tasked to judge the presence or absence of the near-vertical Gabor in the noisy image. We adjusted the signal contrast on a participant-by-participant basis to achieve a d' of approximately 1.5 using data from preliminary experiments. Each session consisted of 100 trials, with inter-trial intervals randomized between 800-1200 ms. Participants completed 15-25 sessions in about 2 hours.

The raw data were published in a previous report on serial dependencies in perception[17]. The re-analysis presented here focuses on a different question (the perceptual and temporal quality of FAs as a proxy for hallucinations) and uses a different analytical approach (HMM-GLMs).

5.4.2 Quantification and Statistical Analysis

Responses were categorized into one of four stimulus-response types: Hits (alarms on stimulus trials), correct rejections (rejections on no-stimulus trials), FAs (alarms on no-stimulus trials), and misses (rejections on stimulus trials). We employed standard logistic and linear regression using R-packages lmer, glmer, and afex (see Supplemental Table S1).

5.4.2.1 Classification images Following previous approaches[17], we filtered each static white noise image with the Gaussian stimulus envelope, applied fast 2-D Fourier-transformation, and extracted the power at the Gabor patch's spatial frequency, separately for orientations from 0 to 180 degrees in three degree intervals. For every image, we computed the average power between 60 to 120 degrees, relative to the overall power from complete orientation range. To visualize the classification images, we subtracted the average power-by-orientation at rejection trials from the average power-by-orientation at alarm trials. To examine serial dependence, the same analysis was conducted separately for trials following clockwise (CW) and counter-clockwise (CCW) inducers. In each condition and participant, we determined the average vector of the classification image by computing the centroid of a polygon composed of vector endpoints in polar coordinates.

5.4.2.2 Hidden Markov Modeling We used General Linear Models (GLM) to predict the response y_t (0: rejection: 1: alarm) from the contrast at the current trial (s_t: zero contrast, low contrast, high contrast) and the response at the preceding trial y_{t-1}:

$$P(y_t = 1|x_t) = \frac{1}{1 + e^{-x_t \times w}} \tag{1}$$

$$x_t \times w = s_t \times \beta_s + y_{t-1} \times \beta_H \tag{2}$$

We then used Hidden Markov Models (HMMs, as implemented in the *SSM: Bayesian learning and inference for state space models* python library, https://github.com/lindermanlab/ssm) to model transitions between two latent states k and j. At each trial, the HMM predicted the latent state z_t was k or j, with alternation between k and j that were governed by a 2 x 2 transition matrix A:

Each latent state was linked to an independent GLM that predicted y_t from s_t and y_{t-1} based

12

on the weights $w_{j/k}$:

$$P(y_t = 1|x_t, z_t) = \frac{1}{1 + e^{-x_t \times w_{k/j}}} \tag{3}$$

$$x_t \times w_{k/j} = s_t \times \beta_{s,k/j} + y_{t-1} \times \beta_{p,k/j} \tag{4}$$

For both states, parameters were initialized at the regression weights of a one-state control GLM. To compare the performance of the two-state GLM relative to the one-state control GLM, we computed the difference in Bayesian Information Criterion (BIC) between the two models. The SSM hyperparameters were defined as follows: $\sigma^2 = 10$ (variance over the GLM weights $w_{k/j}$, and $\alpha = 1$ (Dirichlet prior over the transition matrix A) was set to 1. The latent states k and j were linked to mode by comparing β_s and β_H (external mode: $\beta_s > \beta_H$, internal mode: $\beta_H > \beta_s$). We labelled trial as external when $P(z_t = external) = 1 - P(z_t = internal) > 0.5$.
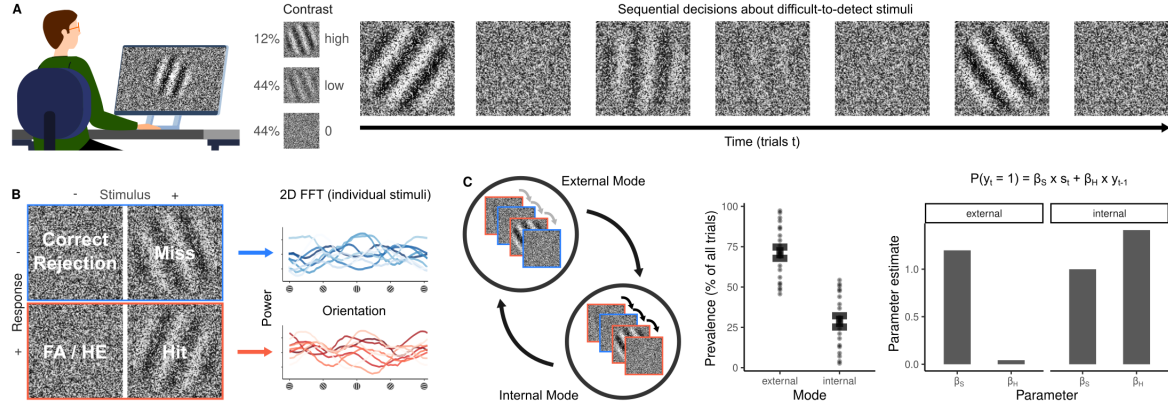
6 Figures

6.1 Figure 1

Figure 1.

A. Paradigm. Participants were instructed to report whether they perceived (close-to-) vertical gratings embedded in noise (left panel). At stimulus trials, gratings were presented either at low or at high contrast. At no-stimulus trials, no grating was presented (contrast = 0). We reasoned that alarms would be more frequent when participant expect to experience a signal, and when the sensory noise is compatible with the predicted signal ($(P(signal|noise) = P(signal) \times P(noise|signal)$, upper panel). We induced signal expectations $P(signal)$ by presenting stimuli in sequence (right panel). We hypothesized that such a sequence would induce false alarms (FAs) via the phenomenon of serial dependence. Serial dependencies cause perception to be biased toward preceding experiences, causing alarms to be more likely to be followed by alarms, and rejections more likely to be followed by rejections. Since serial dependencies are known to be stronger after confident experiences, we predicted that FAs would be more frequent after high-contrast stimuli. In the example shown on the right, our hypothesis therefore predicts that FAs would me most likely at the 2nd and 7th trial.

B. Predictive perceptual templates. To compute the compatibility of the white noise with the expected signal ($P(noise|signal)$), we subtracted the power-by-orientation distribution of the random white noise images at rejection trials (blue) from the power-by-orientation

14

distribution of the random noise images at alarm trials (red). The right panel shows the power-by-orientation distributions of 20 example stimuli that occurred at alarm and rejection trials, respectively.

C. External and internal modes in perception. Perception is known to slowly alternate between two opposing modes (left panel): During the external mode, perception is dominated by incoming stimuli, with weak serial dependencies between subsequent trials (grey arrows). By contrast, during the internal mode, perception is strongly biased by preceding experiences (strong serial dependencies, black arrows). The examples depicted within the circles illustrate the effect of external and internal mode on perception: In the external mode, perception dissociates from predictions that stem from serial dependencies, such that alarms (red) and rejections (blue) are equally likely to occur after high-contrast stimuli. In the internal mode, perception is strongly biased toward previous experiences, causing alarms (red) to be more frequent after high-contrast stimuli. Between-mode alternations can by discovered by Hidden Markov Models (HMM). HMMs describe alternations between two independent General Linear Models (GLMs), each of which predicts perceptual experiences y_t via weights assigned to the stimulus $\beta_S \times s_t$ at the current trial, and the experience made at the preceding trial $\beta_H \times y_{t-1}$. In this experiment, the GLM-HMM discovered an external mode ($71.3 \pm 3.5\%$ percent of trials), during which β_S is high while β_H is low (middle and right panels), and an internal mode ($28.7 \pm 3.5\%$ percent of trials), during which β_S is low while β_H is high.
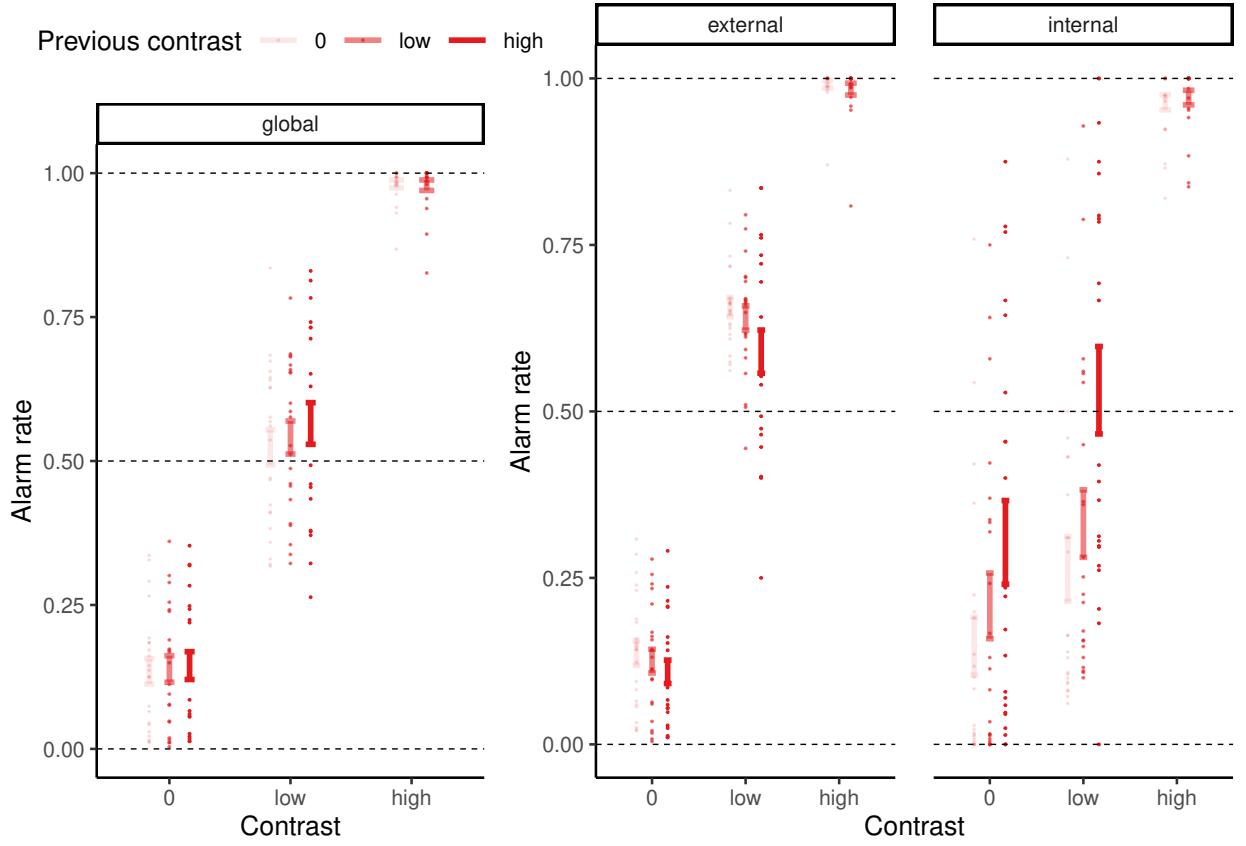
6.2 Figure 2

Figure 2. The effect of serial dependencies on alarms and rejections depend on mode. The global alarm rate correlated positively with the contrast of stimuli presented at the preceding trial (0.11 ± 0.03, z = 4.33, p = 1.5×10^{-5}). This suggests that serial dependencies contribute to FAs. When separating our data in episodes of external and internal modes, we found positive serial dependencies only during the internal mode (0.96 ± 0.06, z = 16.96, p = 1.71×10^{-64}). During the external mode, by contrast, we observed a negative correlation between alarm rate and preceding stimulus contrast (-0.19 ± 0.03, z = -6.31, p = 2.72×10^{-10}).
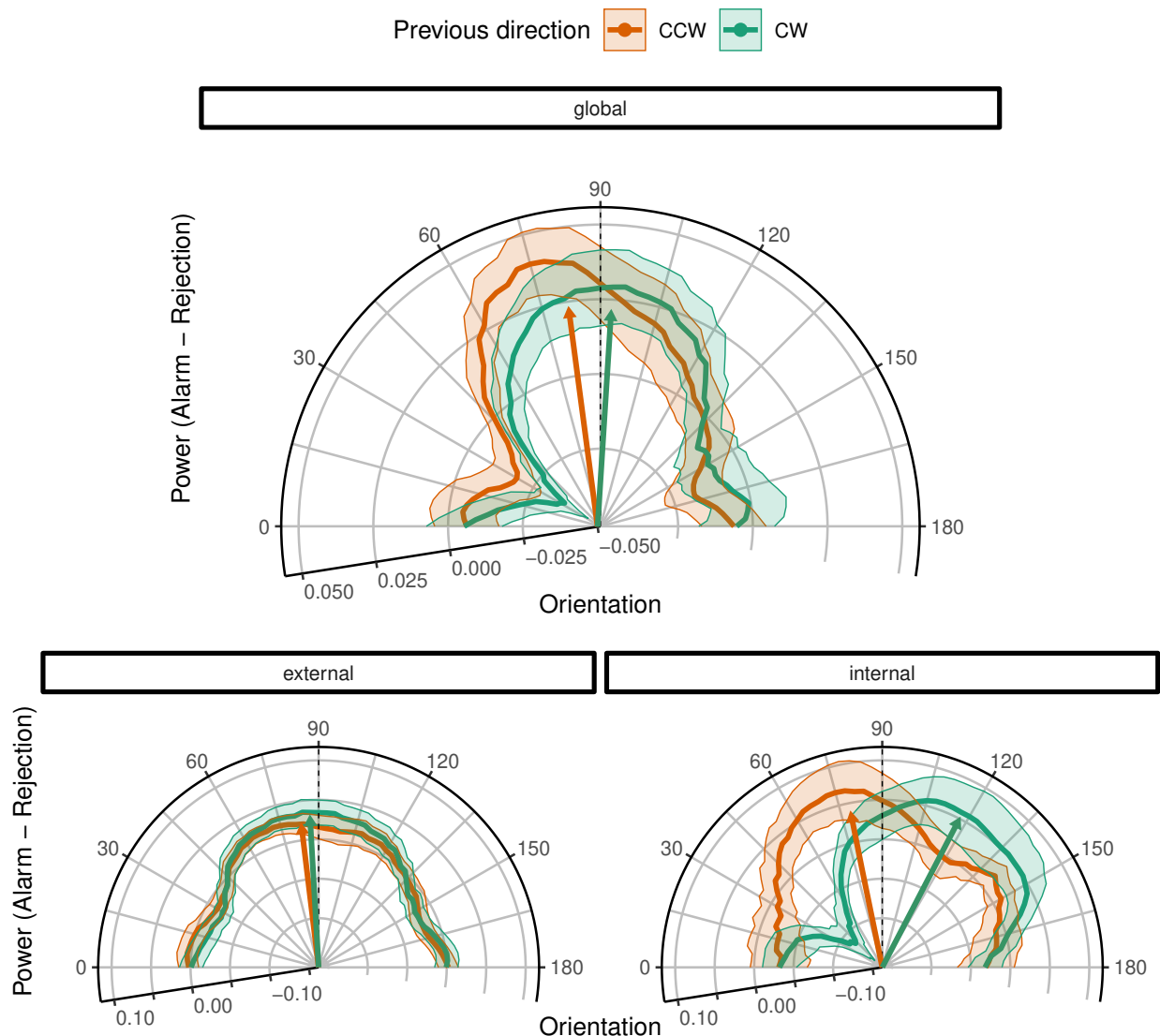
6.3 Figure 3

Figure 3. The effect of predictive templates on alarms and rejections depends on mode. Globally, the classification images computed for alarms (power-by-orientation distribution at alarm trials relative to the power-by-orientation distribution at rejection trials) revealed a peak around 90°. This corresponds to the average orientation of the signal (0.1 ± 0.02, T(1.37×10^4) = 7.24, p = 4.84×10^{-13}). During the internal mode, we found a shift of the close-to-vertical noise power peak toward the orientation of the preceding stimulus (19.87 ± 6.73, T(20.28) = 2.95, p = 7.82×10^{-3}; clockwise tilt [CW] in green, counter-clockwise tilt

328 [CCW] in orange; dots indicate position of the average vector). This shift was not observed

329 during the external mode (-3.67 ± 6.27, T$(20.15) = -0.59$, p $= 0.56$).

7 Supplementary Information

7.1 Supplemental Figure S1
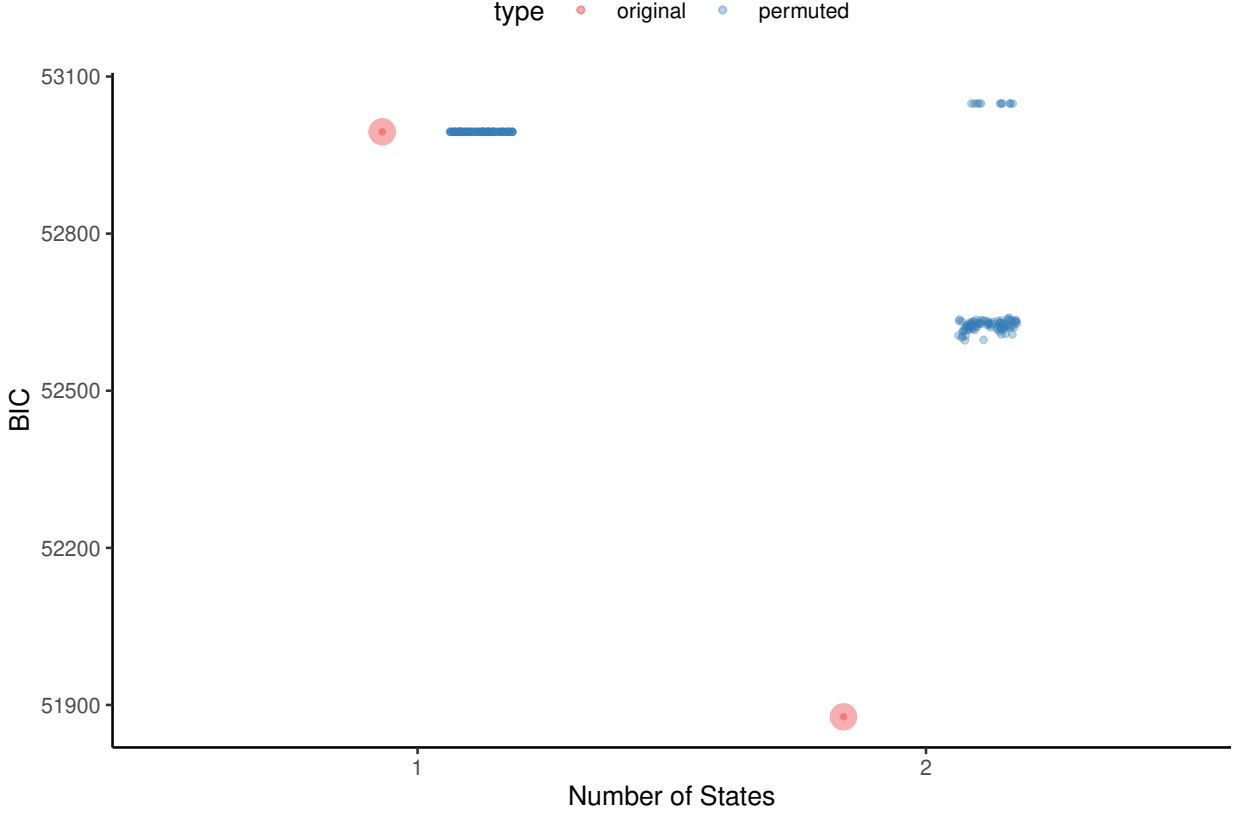
Supplemental Figure S1. We used Bayesian model comparison to evaluate whether the two-state HMM-GLM (number of parameters: 8) provided a more parsimonious explanation of our data in comparison to the one-state control model (number of parameters: 2). To compare the two models, we computed the difference in $\text{BIC} = \log(N) \times k - 2 \times \log(L)$. Relative to the one-state GLM, the two-state HMM-GLM yielded a reduction in BIC by $\delta_{BIC} = -1.12 \times 10^3$ (red). We did not observe this reduction when we randomly shuffled the trials for a total of 100 iterations, and computed BICs for these randomly permuted data (average $\delta_{BIC} = -323.8$; blue). As expected, shuffling did not have an effect on BIC in the one-state control GLM.

7.2 Supplemental Figure S2
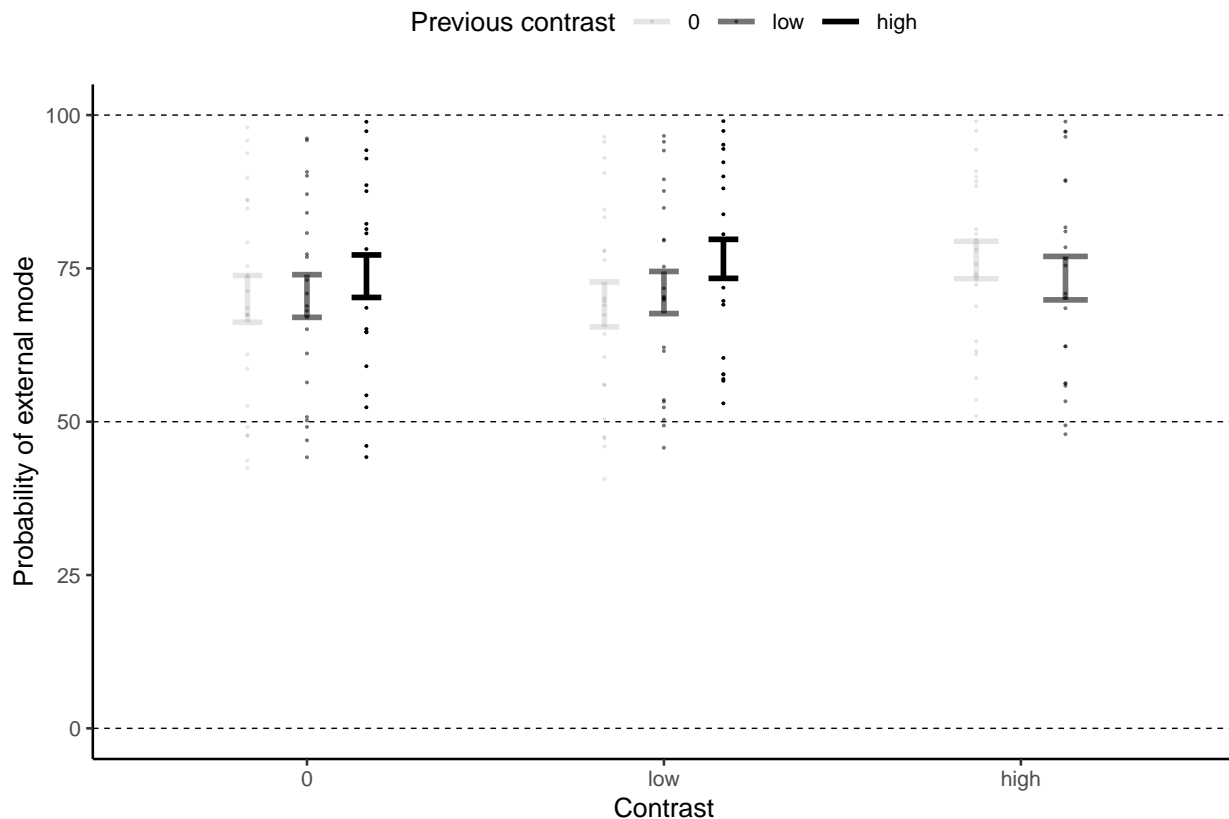
Supplemental Figure S2. As an alternative explanation for external and internal modes, one may assume a response heuristic that uses the external sensory data when it is reliable (i.e., at high-contrast trials), and applies internal predictions when the external sensory data is not reliable (i.e., no-stimulus and low-contrast trials). The latter explanation would entail that the trial-wise probability of external mode depends exclusively on contrast. We found that the probability of the external mode was higher when the contrast at the current trial was high (0.19 ± 0.03, z = 7.58, p = 3.51×10^{-14}). Likewise, the external mode was more likely when the contrast of the stimulus presented at the preceding trial was high (0.22 ± 0.03, z = 8.5, p = 1.82×10^{-17}). This may suggest that stimuli with high signal-to-noise ratio make switches to the external mode more likely. However, the overall change in the probability of external mode was small, ranging form a minimum of 69.12% to a maximum of 76.56%. Moreover, the average interval between mode switches (59.1 ± 8.8 trials) was almost eight times longer than the average interval between high-contrast stimuli (8.66 ± 0.06

20

trials). Our analyses therefore reveal a potential role of high signal-to-noise ratio for mode alternations. They are, however, not compatible with the view that external and internal modes reflect response strategies applied exclusively to stimuli at high and low signal-to-noise ratio, respectively.

7.3 Supplemental Table S1

RESOURCE	SOURCE	IDENTIFIER
Deposited data & code		
Analyzed data & custom code	https://github.com/veithweilnhammer/predictive_templates/	N/A
Software		
Python 3	http://www.python.org/	RRID:SCR_008394
Jupyter Notebook	https://jupyter.org/	RRID:SCR_018315
numpy	http://www.numpy.org	RRID:SCR_008633
pandas	https://pandas.pydata.org	RRID:SCR_018214
SSM	https://github.com/lindermanlab/ssm	N/A
Matlab	https://www.mathworks.com/	RRID:SCR_001622
Psychtoolbox 3	http://psychtoolbox.org/	RRID:SCR_002881
R	http://www.r-project.org/	RRID:SCR_001905
RStudio	https://www.rstudio.com/	RRID:SCR_000432
lme4, afex, ggplot2, ggridges, gridExtra, tidyr, plyr	http://cran.r-project.org/	RRID:SCR_003005

References

1. Powers, A. R. *et al.* Pavlovian conditioning–induced hallucinations result from over-weighting of perceptual priors. *Science* **357**, 596–600 (2017).

2. Stuke, H. *et al.* Overly Strong Priors for Socially Meaningful Visual Signals Are Linked to Psychosis Proneness in Healthy Individuals. *Frontiers in Psychology* **12**, (2021).

3. 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).

4. Kafadar, E. *et al.* Conditioned Hallucinations and Prior Overweighting Are State-Sensitive Markers of Hallucination Susceptibility. *Biological Psychiatry* **92**, 772–780 (2022).

5. Leptourgos, P. *et al.* Relating Glutamate, Conditioned, and Clinical Hallucinations via 1H-MR Spectroscopy. *Schizophrenia Bulletin* **48**, 912–920 (2022).

6. Horga, G. *et al.* An integrative framework for perceptual disturbances in psychosis. *Nature Reviews Neuroscience* **20**, 763–778 (2019).

7. Schmack, K. *et al.* Striatal dopamine mediates hallucination-like perception in mice. *Science (New York, N.Y.)* **372**, eabf4740 (2021).

8. Sterzer, P. *et al.* The Predictive Coding Account of Psychosis. *Biological Psychiatry* **84**, 634–643 (2018).

9. Corlett, P. R. *et al.* Hallucinations and Strong Priors. *Tics* **23**, 114–127 (2019).

10. Howes, O. D. *et al.* The dopamine hypothesis of schizophrenia: Version III - The final common pathway. *Schizophrenia Bulletin* **35**, 549–562 (2009).

11. Corlett, P. R. *et al.* Prediction error, ketamine and psychosis: An updated model. *Journal of Psychopharmacology* **30**, 1145–1155 (2016).

12. Toh, W. L. *et al.* On the phenomenology of auditory verbal hallucinations in affective and non-affective psychosis. *Psychiatry Research* **290**, 113147 (2020).

13. López-Silva, P. *et al.* The phenomenology of auditory verbal hallucinations in schizophrenia and the challenge from pseudohallucinations. *Frontiers in Psychiatry* **13**, (2022).

14. Parnas, J. *et al.* Phenomenology of auditory verbal hallucination in schizophrenia: An erroneous perception or something else? *Schizophrenia Research* **265**, 83–88 (2024).

15. Oorschot, M. *et al.* Temporal dynamics of visual and auditory hallucinations in psychosis. *Schizophrenia Research* **140**, 77–82 (2012).

16. Hermans, K. *et al.* Temporal dynamics of suspiciousness and hallucinations in clinical high risk and first episode psychosis. *Psychiatry Research* **290**, 113039 (2020).

17. Murai, Y. *et al.* Serial dependence revealed in history-dependent perceptual templates. *Current biology: CB* **31**, 3185–3191.e3 (2021).

18. Weilnhammer, V. *et al.* N-Methyl-D-aspartate receptor hypofunction causes recurrent and transient failures of perceptual inference. (2024) doi:10.1101/2024.05.24.595590.

19. Weilnhammer, V. *et al.* Bistable perception alternates between internal and external modes of sensory processing. *iScience* **24**, 102234 (2021).

20. Weilnhammer, V. *et al.* Sensory processing in humans and mice fluctuates between external and internal modes. *PLOS Biology* **21**, e3002410 (2023).

21. Manassi, M. *et al.* Continuity fields enhance visual perception through positive serial dependence. *Nature Reviews Psychology* **3**, 352–366 (2024).

22. Fischer, J. *et al.* Serial dependence in visual perception. *Nat. Neurosci.* **17**, 738–743 (2014).

23. Kiyonaga, A. *et al.* Serial Dependence across Perception, Attention, and Memory. *Trends in Cognitive Sciences* **21**, 493–497 (2017).

24. Pascucci, D. *et al.* Serial dependence in visual perception: A review. *Journal of Vision* **23**, 9 (2023).

25. Cicchini, G. M. *et al.* Serial Dependence in Perception. *Annual Review of Psychology* **75**, 129–154 (2024).

26. Manassi, M. *et al.* Serial dependence in visual perception: A meta-analysis and review. *Journal of Vision* **23**, 18 (2023).

27. Cicchini, G. M. *et al.* The functional role of serial dependence. *Proceedings. Biological Sciences* **285**, 20181722 (2018).

28. Ashwood, Z. C. *et al.* Mice alternate between discrete strategies during perceptual decision-making. *Nature Neuroscience* **25**, 201–212 (2022).

29. Eckert, A.-L. *et al.* Cross-Modality Evidence for Reduced Choice History Biases in Psychosis-Prone Individuals. *Schizophrenia Bulletin* **49**, 397–406 (2023).

30. Stein, H. *et al.* Reduced serial dependence suggests deficits in synaptic potentiation in anti-NMDAR encephalitis and schizophrenia. *Nature Communications* **11**, 1–11 (2020).