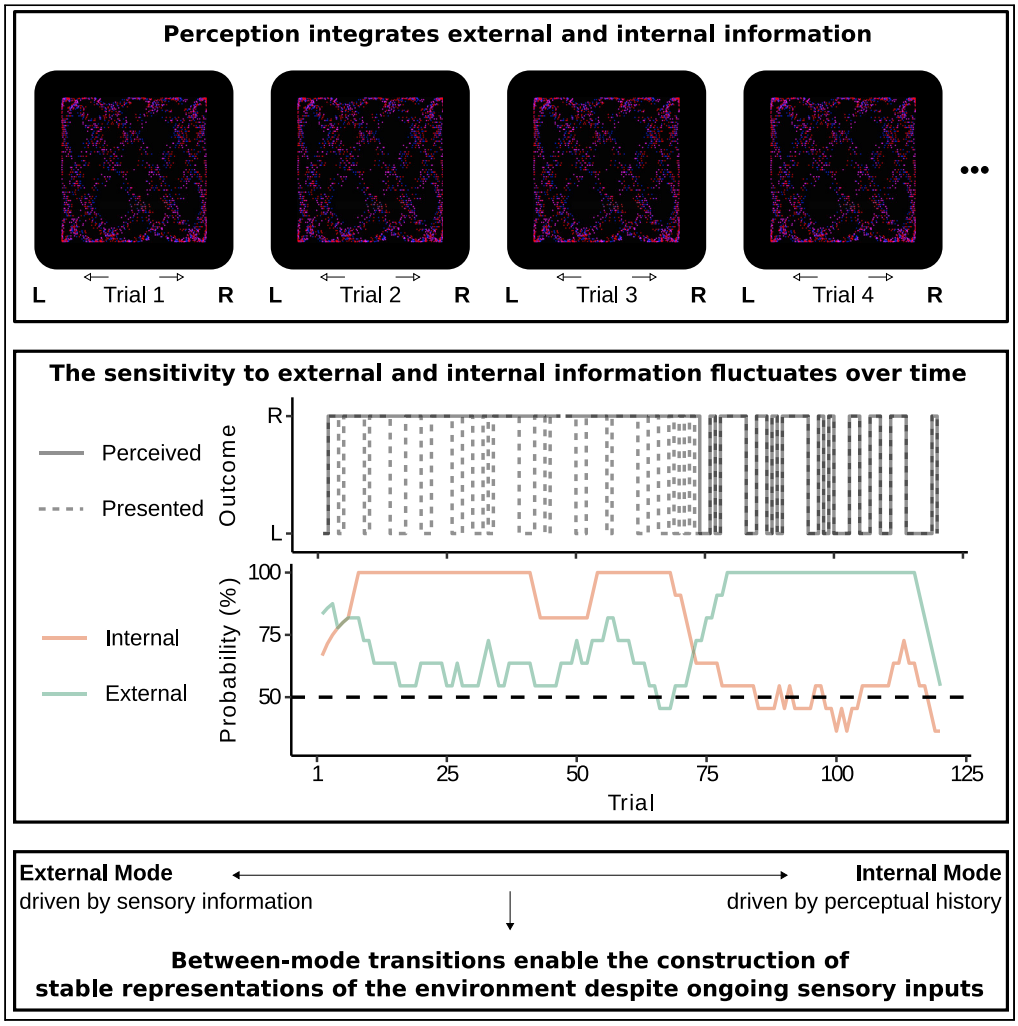
iScience

CellPress
OPEN ACCESS

Article

Bistable perception alternates between internal and external modes of sensory processing



Veith Weilnhammer, Meera Chikermane, Philipp Sterzer

veith-andreas.weilnhammer@charite.de,
https://osf.io/y2cfm/

HIGHLIGHTS

Perception fluctuates between external and internal modes of sensory processing

Weilnhammer et al., iScience *24*, 102234
March 19, 2021
https://doi.org/10.1016/j.isci.2021.102234

iScience

CellPress
OPEN ACCESS

Article

Bistable perception alternates between internal and external modes of sensory processing

Veith Weilnhammer,[1,2,5,*] Meera Chikermane,[1] and Philipp Sterzer[1,2,3,4]

SUMMARY

Perceptual history can exert pronounced effects on the contents of conscious experience: when confronted with completely ambiguous stimuli, perception does not waver at random between diverging stimulus interpretations but sticks with recent percepts for prolonged intervals. Here, we investigated the relevance of perceptual history in situations more similar to everyday experience, where sensory stimuli are usually not completely ambiguous. Using partially ambiguous visual stimuli, we found that the balance between past and present is not stable over time but slowly fluctuates between two opposing modes. For time periods of up to several minutes, perception was either largely determined by perceptual history or driven predominantly by disambiguating sensory evidence. Computational modeling suggested that the construction of unambiguous conscious experiences is modulated by slow fluctuations between internally and externally oriented modes of sensory processing.

INTRODUCTION

Imagine walking down a dark and unfamiliar street. As you struggle to identify potential obstacles, you are confronted with an ongoing stream of sensory signals, each compatible with multiple interpretations. In such situations, your previous perceptual experiences may provide valuable clues about how to interpret the ambiguous sensory data. Yet, relying too heavily on the past is risky, as you may end up overlooking unexpected changes in the environment.

Experimentally, the influence of preceding experiences on perception is usually investigated in tasks that require participants to perform perceptual decisions in a sequence of consecutive trials (Bergen and Jehee, 2019; Fründ et al., 2014). Such experiments reveal that, even in the absence of any correlation between the stimuli that are presented on successive trials, perception is significantly biased toward preceding choices (Abrahamyan et al., 2016; Fischer and Whitney, 2014; Fritsche et al., 2017; Hsu and Wu, 2020; Liberman et al., 2014; Urai et al., 2017, Urai et al., 2019). Importantly, *perceptual history* effects increase when sensory information becomes unreliable (Bergen and Jehee, 2019; Fründ et al., 2014). This reflects the idea that, when making perceptual decisions in situations of uncertainty, the brain may rely more strongly on internal predictions (Friston, 2005, 2010) that reflect the continuity of the sensory environment.

Integrating the *internal* information provided by perceptual history with the available *external* stimulus information may thus benefit perception by preventing erratic responses to unreliable sensory signals (Friston, 2005, 2010). However, the effects of perceptual history may also become mal-adaptive: when relying too strongly on preceding experiences, observers may become prone to ignore conflicting stimuli, which may lead to *hallucinatory* perceptual states that diverge from the true cause of the sensory data (Horga and Abi-Dargham, 2019; Powers et al., 2017).

In this work, we studied how visual perception balances external with internal sources of information in situations where perceptual history has a particularly strong effect. To this end, we investigated how preceding experiences impact the perception of ambiguous stimuli, i.e., stimuli that are compatible with two mutually exclusive perceptual states and typically give rise to bistable perception (Leopold et al., 2002). During bistable perception, observers experience spontaneous transitions between the two perceptual states, whereas the sensory data remain constant (Logothetis et al., 1996). Importantly, when the ambiguous stimuli are presented in successive trials separated by blank intervals, perception tends to stabilize

[1]Department of Psychiatry, Charité-Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Charitéplatz 1, 10117 Berlin, Germany

[2]Berlin Institute of Health, Charité-Universitätsmedizin Berlin and Max Delbrück Center, 10178 Berlin, Germany

[3]Bernstein Center for Computational Neuroscience, Charité-Universitätsmedizin Berlin, 10117 Berlin, Germany

[4]Berlin School of Mind and Brain, Humboldt-Universität zu Berlin, 10099 Berlin, Germany

[5]Lead contact

*Correspondence: veith-andreas.weilnhammer@charite.de, https://osf.io/y2cfm/

https://doi.org/10.1016/j.isci.2021.102234

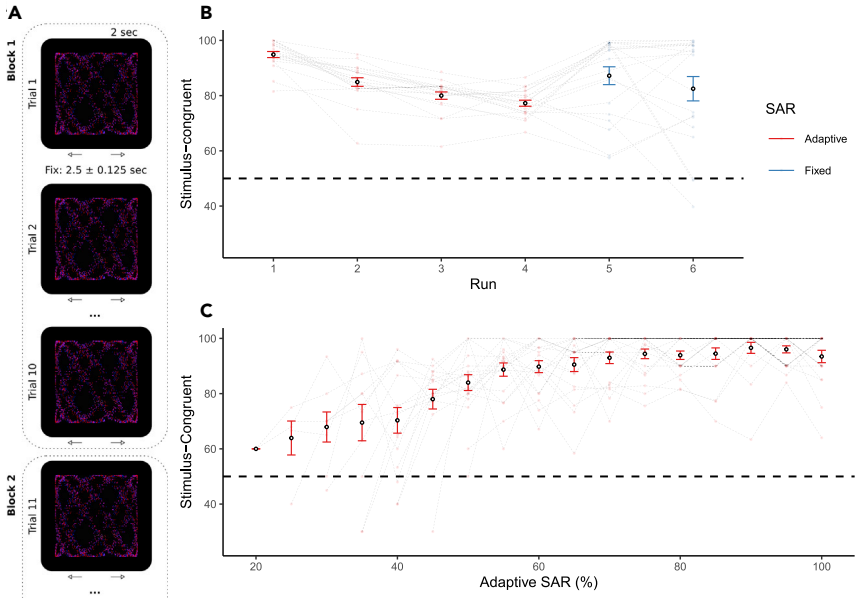


Figure 1. Psychophysical staircase

(A) **Graded ambiguity**. Participants viewed partially ambiguous structure-from-motion stimuli and indicated whether they perceived 3D rotation to the left or to the right. In runs R1-4, we dynamically adjusted the signal-to-ambiguity ratio (SAR) according to a staircase procedure that was based on the number of stimulus-congruent trials computed within blocks of 10 successive trials. During the final runs R5 and R6, we fixed the SAR to the average SAR obtained during runs R1-4.

(B) **Stimulus-congruent percepts across runs.** In runs R1-4 (depicted in red), the staircase procedure introduced dynamic adjustments in the SAR, reducing the frequency of stimulus-congruent percepts to approximately 75% (R1: 94.88 ± 1.1%; R2: 84.92 ± 1.55%; R3: 80 ± 1.33%; R4: 77.25 ± 1.09%). In runs R5-6 (depicted in blue), the SAR was fixed to the average SAR from the preceding runs R1-4 (60.25 ± 2.36%). Stimulus-congruent percepts amounted to 87.21 ± 3.23% in R5 and 82.5 ± 4.41% in R6.

(C) **Stimulus-congruent percepts across levels of SAR.** Stimulus-congruent percepts were more frequent at higher levels of disambiguating sensory information, ceiling at 100%. Pooled data are represented as mean ± SEM.

in one of the two interpretations (Maloney et al., 2005), indicating a pronounced effect of perceptual history (Pearson and Brascamp, 2008).

Here, we estimated the strength of perceptual history during bistable perception using a staircase procedure that dynamically adjusted the degree of perceptual ambiguity of structure-from-motion stimuli. By quantifying the effect of perceptual history relative to graded levels of sensory ambiguity, we investigated the computational mechanisms of integrating internal with external information during bistable perception.

RESULTS

To study how perceptual history is balanced against external sensory information during bistable perception, we asked 20 participants to indicate whether they perceived partially ambiguous random-dot-kinematograms as rotating to the left or the right (Figure 1A and Video S1). At each trial, we attached a 3D signal to a subset of the stimulus dots. This enabled us to parametrically manipulate the stimulus' *signal-to-ambiguity ratio* (Weilnhammer et al., 2020) (SAR). Ranging between 0% and 100%, these varying *levels* of disambiguating sensory information enforced one of the two stimulus interpretations (i.e., the *direction* of disambiguation). Within each experimental run, both directions of disambiguation occurred in equal number and in random sequence.

Perception integrates perceptual history with disambiguating sensory information

In the first four runs (R1-4, Figure 1B), we estimated individual threshold SARs necessary to induce balanced frequencies of stimulus-congruent and stimulus-incongruent percepts (i.e., trials perceived as congruent or incongruent with the disambiguating sensory information, respectively; Figure 2A). To this end, we dynamically adjusted the SAR based on the proportion of stimulus-congruent responses in consecutive 10-trial blocks. This psychophysical staircase decreased the SAR if less than 80% of trials were perceived as stimulus-congruent.

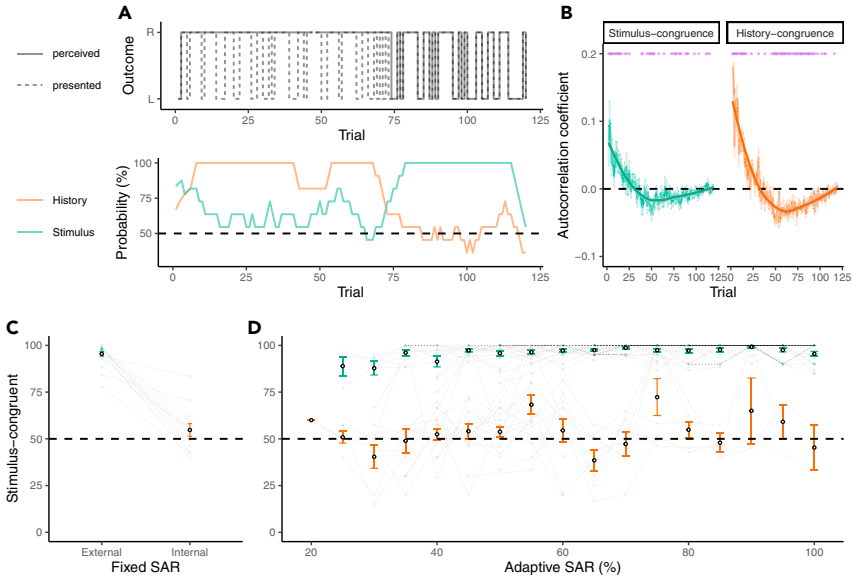


Figure 2. External and internal modes

(A) **Stimulus- and history-congruent perceptual states**. To visualize the influence of disambiguating sensory information and perceptual history, the upper panel depicts the time course of presented stimuli (L/R: disambiguating stimulus information for leftward/rightward rotation; dashed line) and the associated time course of perception (solid line). Perception is stimulus-congruent when the presented stimulus matches the associated perceptual state (i.e., overlap between the dashed and the solid line). History-congruent perception occurs when the perceptual state at a given trial matches the perceptual state at the preceding trial. The lower panel depicts the dynamic probabilities of stimulus-congruent percepts (green) and history-congruent percepts (orange) computed in sliding windows of ±5 trials for a representative participant. Perceptual processing switched between prolonged intervals of *internal mode* (green line below orange line), *external mode* (green line above orange line), and *intermediate mode* (overlap between green an orange line).

(B) **Average autocorrelation coefficients of stimulus- and history-congruence.** Despite constant SAR at threshold, both stimulus and history congruent were highly autocorrelated. If the index trial was perceived as congruent with visual stimulation (left panel) or perceptual history (right panel), the observer was more likely to experience stimulus- or history-congruent perceptual states, respectively, for approximately 25 trials. After that, the observer was more likely to experience incongruent states. The opposite relation holds for incongruent perceptual states at the index trial. Group-level averages were fitted using local polynomial regression fitting. Purple dots indicate trials at which the autocorrelation coefficients differed significantly from chance level ($p < 0.05$, two-sided one-sample t tests).

(C) **Stimulus-congruent percepts during internal and external mode for SAR at threshold.** During external mode, stimulus-congruent percepts made up for almost 100% of trials (95.49 ± 1.28%) but, interestingly, did not differ significantly from chance level during internal mode (54.71 ± 3.39%).

(D) **Stimulus-congruent percepts during internal and external mode across the full range of SAR.** Linear mixed effects modeling indicated that the frequency of stimulus-congruent percepts increased with levels of *SAR*. Internal mode was associated with a strong reduction of stimulus-congruent percepts (main effect of *mode*), which was more pronounced at low levels of SAR (*mode* × *SAR* interaction). Please note that any main effect of *mode* was expected, because external and internal mode were defined based on the dynamic probability of stimulus congruence. Pooled data are represented as mean ± SEM.

Conversely, we increased the SAR if the proportion of stimulus-congruent trials fell below 80%. As expected, stimulus-congruent percepts were less frequent at lower SARs ($F(1, 265.07) = 181.5$, $p = 7.25 \times 10^{-32}$, $BF_{10} = 5.22 \times 10^{28}$, main effect of *SAR*, Figure 1C). In runs R5-6, stimuli were presented at the individual threshold SAR (i.e., the average SAR from runs R1-4), which yielded stimulus-congruent percepts in 84.85 ± 3.12% of trials (Figure 1B).

Conversely, higher SARs reduced the impact of perceptual history (Figure S1). This resulted in a strong inverse relationship between stimulus- and history-congruent percepts (i.e., trials perceived in congruence with the immediately preceding percept), which were anti-correlated both within (average Pearson correlation coefficient $\rho = -0.9 \pm 0.02$, $T(19) = -49.25$, $p = 1.66 \times 10^{-21}$, $BF_{10} = 1.34 \times 10^{18}$, one-sample t test; Figure S2A) and across participants ($\rho = -0.77$, $p = 7.2 \times 10^{-5}$, $BF_{10} = 203.27$, Pearson correlation; Figure S2B).

We did not find any systematic bias toward one of the two perceptual interpretations (average probability of rightward rotation: 51.86 ± 3.04%; T(19) = 0.61, p = 0.55, BF_{10} = 0.27, one-sample t test). Absolute biases were small, amounting to 13.99 ± 1.57% across participants. Error responses were negligible, occurring in only 1.6 ± 1.57% of trials. Unclear percepts were not reported by the participants.

In logistic regression applied to each individual participant's behavioral data, trial-wise perceptual responses were best predicted based on both the current sensory information and the previous percept, as compared with reduced logistic regression models (Figure S2C) that used only stimulus information (T(19) = −9.39, p = 1.45×10^{-8}, BF_{10} = 8.89×10^{5}, paired t test) or only perceptual history (T(19) = −16.46, p = 1.06×10^{-12}, BF_{10} = 6.54×10^{9}) for prediction.

Two additional control analyses confirmed that both disambiguating sensory information and perceptual history significantly modulated the perception of partially disambiguated stimuli. Firstly, general linear mixed effects modeling with a binomial link function indicated a highly significant effect of both disambiguating sensory evidence (z = 45.55; p = 0) and perceptual history (z = 28.51; p = 8.62×10^{-179}), while controlling for the within-participant correlations using random intercepts.

A second possibility for this group-level inference is provided by general estimating equations (Hanley, 2003), which offer a non-parametric way of accounting for within-participant correlation by estimating population average effects. Likewise, this approach revealed a highly significant effect of disambiguating sensory evidence (Wald = 38.6; p = 5.2×10^{-10}) and perceptual history (Wald = 74.33; p = 0, correlation structure = "independence").

These results indicate that the effect of perceptual history is not limited to fully ambiguous stimuli (Pearson and Brascamp, 2008) but modulates perception through a weighted integration with varying levels of disambiguating sensory information (Bergen and Jehee, 2019). This finding aligns with the well-known observation that perception is co-determined by both sensory data and past experiences (Chopin and Mamassian, 2012; Fischer and Whitney, 2014; Fritsche et al., 2017; Hsu and Wu, 2020; Liberman et al., 2014). Perceptual history may benefit perception as an internal representation (Friston, 2005, 2010; Körding and Wolpert, 2004; Teufel and Fletcher, 2020) that stabilizes conscious experience when external sensory information is incomplete or unreliable. On your night-time walks, previous experiences may thus help you to avoid responding to irrelevant fluctuations in the ongoing stream of ambiguous sensory signals.

Perception fluctuates between temporally extended modes that are biased toward either external or internal information

In a next step, we examined how the probabilities of stimulus- and history-congruent percepts evolved within individual runs of the experiment (Figure 2A). Intriguingly, we found that both stimulus- and history-congruence were significantly autocorrelated (Figure 2B), indicating that the integration of perceptual history with sensory information was highly variable over time. For partially ambiguous stimuli presented at constant SARs (R5-6), we observed marked switches between intervals in which perception was either strongly driven by disambiguating sensory information (*external mode*, 73.25 ± 6.17% of trials) or determined by perceptual history (*internal mode*; 23.94 ± 5.84%), in addition to shorter intermediate intervals (2.81 ± 0.77%; Figure 2A, lower panel). Switches between these modes occurred on average every 39.9 ± 7.31 trials (179.53 ± 32.91 s).

Our analyses therefore revealed prolonged intervals of alternating biases toward either internal or external information. This finding is incompatible with the view that perception is best explained by integrating uncertain sensory data with only the immediately preceding perceptual state. As indicated by simulation analyses (Figure S3), such a Markovian assumption did not reproduce the autocorrelation of stimulus and history congruence (Figure S3B) and predicted longer external (T(19) = 2.75, p = 0.01, BF_{10} = 4.17, paired t test; Figure S3C) as well as shorter internal modes (T(19) = −3.49, p = 2.44×10^{-3}, BF_{10} = 16.92).

In sum, these results imply that a stable moment-by-moment integration of current sensory information with the immediately preceding percept is not sufficient to explain the perceptual dynamics during graded ambiguity. Rather, our findings suggest that participants transition between temporally extended perceptual *modes* (Honey et al., 2017) that are biased toward either external information (i.e., disambiguating sensory information) or internal information (i.e., perceptual history).

Importantly, switches between internal and external modes could not be attributed to small fluctuations in the participants' sensitivity to disambiguating sensory information. At threshold (R5-6), stimulus-congruent percepts were close to 100% during external mode but ranged at chance level during internal mode ($T(12) = 1.39$, $p = 0.19$, $BF_{10} = 0.61$, one-sample t test, Figure 2C). Please note that the overall difference in stimulus-congruency between modes is expected, because external and internal mode were defined based on the dynamic probability of stimulus-congruent perceptual states.

Moreover, internal mode suppressed the sensitivity to disambiguating sensory information not only at the threshold but across the full range of SAR ($F(2, 484.41) = 35.26$, $p = 5.04 \times 10^{-15}$, $BF_{10} = 4.78 \times 10^{66}$; main effect of *mode*; Figure 2D). During runs in which the SAR was adjusted dynamically (R1-R4), transitions from internal to mode were more likely to occur when the available sensory information was reduced ($F(2, 472.71) = 5.25$, $p = 5.58 \times 10^{-3}$, $BF_{10} = 3.57$, *mode* × *SAR* interaction). In sum, these control analyses argue against the view that between-mode transition may result exclusively from a threshold phenomenon.

As a second caveat, we asked whether the observed transitions between internal and external mode constitute a perceptual phenomenon or, alternatively, occur only due to cognitive processes that are situated downstream of perception (Brascamp et al., 2018). In this context, it may be argued that the participants' attention to the experimental task may have fluctuated over time (Rosenberg et al., 2013; Zalta et al., 2020), leading to intervals of stereotypic reporting behavior. We addressed this potential confound by analyzing response times (*RTs*, see Figure S4), which have been shown to link closely with on-task attention (Prado et al., 2011; Rosenberg et al., 2013).

In contrast to stimulus and history congruence, response times remained stable across the experimental runs (Figure S4A) and did not vary across levels of SAR (R1-R4; $F(1, 261.5) = 0.05$, $p = 0.82$, $BF_{10} = 0.15$; Figure S4B). At threshold, RTs did not differ between external and internal mode ($T(12) = 0.74$, $p = 0.48$, $BF_{10} = 0.35$, paired t test; Figure S4C). Moreover, when analyzing RTs according to the factors *mode* and *SAR* in runs R1-R4 (Figure S4D), we found that, during internal mode, RTs increased for escalating levels of SAR. Speculatively, this *mode* × *SAR* interaction ($F(2, 476.5) = 10.73$, $p = 2.77 \times 10^{-5}$, $BF_{10} = 538.42$) could reflect the increase in conflict between the history-congruent state and the available sensory information (Weilnhammer et al., 2020).

At the same time, both the absence of any mode effect on RTs at threshold as well as the sensitivity of internal-mode RTs to levels of *SAR* argue against the notion that internal mode is caused by the participants paying less attention to the experimental task. In line with this observation, we found no changes in the distribution of normalized RTs, as participants transitioned between internal and external mode (see Figure 4E for group RTs collapsed across participants and Figure 4F for individual distributions).

As a final control analysis, we checked whether internal mode was associated with an enhanced impact of the perceptual state *experienced* at the preceding trial (i.e., perceptual history), as opposed to the disambiguating sensory information *presented* at the preceding trial (i.e., stimulus history). As expected, history-congruent perceptual states dominated periods of internal mode processing ($94.15 \pm 1.03\%$), whereas stimulus history had no detectable influence on perception in these intervals ($49.79 \pm 1.03\%$; $T(19) = -0.2$, $p = 0.84$, $BF_{10} = 0.24$, one-sample t test).

During internal mode, perceptual history thus strongly determines conscious experience, overriding otherwise effective sensory information. As you interpret ambiguous sensory information on your walk through the dark, relying on an internal representation of your surroundings may dramatically increase the energy efficiency of perception. However, this is only adaptive in stable environments, i.e., when sensory events are highly auto-correlated. In volatile environments, internally biased sensory processing may cause perception to get stuck in the past, resulting in hallucinatory experiences that ignore relevant conflicts (Weilnhammer et al., 2020) with sensory information (Horga and Abi-Dargham, 2019).

Computational modeling indicates that between-mode transitions are best explained by a fluctuating impact of accumulating perceptual history

How can perception achieve an adaptive balance between external and internal mode? To address this question, we investigated the potential computational mechanisms that could lead to the observed oscillations between internally and externally biased modes of perceptual processing. To this end, we constructed a set of four generative behavioral models (Wilson and Collins, 2019) (Figure 3) that differed across two dimensions.

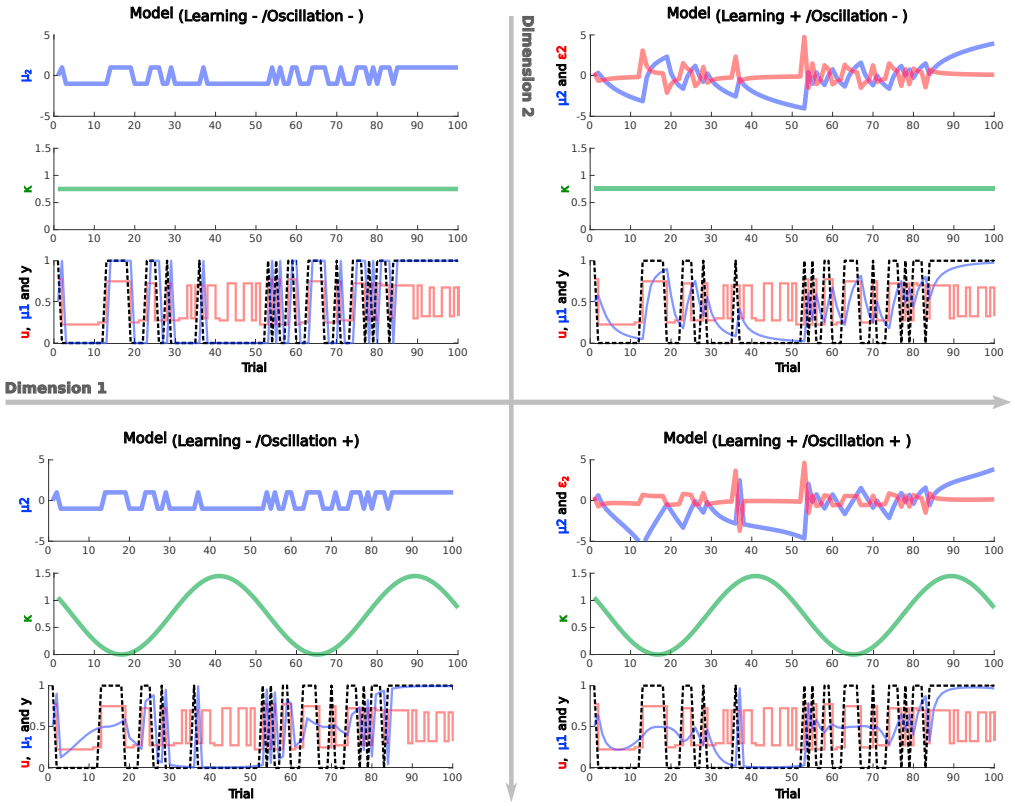


Figure 3. Computational modeling: Modelspace

To investigate the computational mechanisms of between-mode transitions, we constructed a space of four behavioral models that differed along two dimensions. Each model's quantities are shown in three separate panels. Along a **first dimension** (horizontal arrow), we manipulated whether perceptual history effects were represented exclusively by the perceptual state at the preceding trial (left side) or, alternatively, dynamically accumulated according to a learning rate ω (right). Perceptual history and its updating are displayed in the upper panel of each model. The blue line represents μ_2, i.e., the tendency to expect rightward (above zero) or leftward (below zero) rotation at the upcoming trial. The red line depicts dynamic precision-weighted prediction errors ε_2 that update μ_2 in response to the sequence of perceptual experiences. Along a **second dimension** (vertical arrow), we contrasted models that assumed a stable influence of perceptual history on perception (top) against models that assumed a systematic fluctuation in the impact of perceptual history. κ (green line; middle panel) represents the weight at which perceptual history impacts on perception. The lower panel shows the perceptual prediction $\widehat{\mu_2}$ (blue line, provided by a sigmoid transform of $\mu * \kappa$), the disambiguating sensory information *u* (red) and the participants' response *y* (black).

On the first dimension, we asked whether biases toward internal mode arise from the sequence of previous experiences. We reasoned that, if perceptual history effects dynamically accumulate over time (Brascamp et al., 2008; Pearson and Brascamp, 2008), perception would be more strongly biased toward a perceptual state if the current trial was preceded by a long sequence of history-congruent trials. Accumulating perceptual history effects could eventually become strong enough to override otherwise effective sensory information, thereby creating intervals during which perception is strongly determined by internal information.

To this end, we adopted a Bayesian modeling approach that frames perception as an inferential process in which perceptual decisions are determined by posterior distributions (Friston, 2010). Following Bayes' rule, such posterior distributions are computed by integrating a likelihood distribution representing the sensory evidence (i.e., disambiguating sensory information for left- or rightward rotation at a given SAR) with the prior probability of perceptual states (i.e., perceptual history).

The null model $M_{Learning-/Oscillation-}$ (see transparent method section and Figure 3 for details) assumes that the effect of perceptual history (i.e., the estimated prior probability of perceptual states) depends only on

the perceptual state at the immediately preceding trial. Its weight on perception is determined by the parameter κ. The impact of sensory information, in turn, depends on the sensitivity parameter α.

By contrast, in the alternative model $M_{Learning+/Oscillation-}$, the estimated prior probability of perceptual states depends not only on the response at the preceding trial but dynamically accumulates over time according to a two-level Hierarchical Gaussian Filter (Mathys et al., 2014). Thus, the implicit *belief* in the probability of perceiving leftward rotation increases as a function of the number of preceding trials that have been experienced as rotating toward the left (and vice versa). The second-level accumulation of perceptual history is governed by the learning-rate parameter ω.

In this model, switches between modes can only be driven by experience. Once perceptual history effects have accumulated and caused the estimated probability of leftward rotation to increase significantly above chance level, switches to external mode are enabled by prediction errors that are caused by the experience of rightward rotation (and vice versa).

As an alternative explanation, we reasoned that switches between modes could additionally be facilitated by systematic fluctuations in κ, the parameter governing the impact of perceptual history on perception. When κ is low, perceptual states are more likely to be history incongruent, increasing the likelihood of prediction errors that enable the transition from internal to external mode. To test whether such fluctuations provide a plausible explanation of our behavioral data, we introduced a second dimension to our model space by constructing $M_{Learning+/Oscillation+}$ and $M_{Learning-/Oscillation+}$. Instead of estimating κ as a stable parameter, these models enable oscillations in κ that are governed by parameters for amplitude *amp*, frequency *f* (in *nb trials*$^{-1}$), and phase *p*.

We inverted all models based the trial-wise perceptual responses given by our participants and used random-effects Bayesian model family selection (Stephan et al., 2009) to determine whether the dynamic accumulation of perceptual history (dimension 1) and systematic fluctuations in its impact (dimension 2) were likely to represent a computational mechanism of mode switches.

On the first dimension, we found that models assuming a dynamic accumulation of perceptual history (*Learning+*) outperformed *Learning−* models at a protected exceedance probability of 100%. On the second dimension, Bayesian model selection indicated that our data were better explained by models that assumed a fluctuating impact of perceptual history (*Oscillation+*) as compared with *Oscillation−* models at a protected exceedance probability of 99.98%. $M_{Learning+/Oscillation+}$ was therefore identified as the clear winning model (protected exceedance probability = 99.82%; see Figure 4A for model-level inference at the participant level and Figure 4B for posterior parameter estimates).

With this, our computational approach suggests that switches between internal and external mode are governed by two interlinked processes: In line with previous findings (Brascamp et al., 2008; Pearson and Brascamp, 2008), we found that perceptual history accumulates over time. Eventually, accumulating perceptual history may override disambiguating sensory information, causing a transition to from external to internal mode. In isolation, however, such a process falls short of explaining transitions in the opposite direction. Because perceptual history effects continue to accumulate during internal mode, they should eventually become impossible to overcome (Wexler et al., 2015). Crucially, our modeling results propose that fluctuations in the impact of perceptual history enable transition from internal to external mode by temporarily de-coupling the perceptual decision from implicit internal representations of the environment.

DISCUSSION

In this work, we show that perceptual history modulates perception through a weighted integration (Bergen and Jehee, 2019) with varying levels of sensory information. Perceptual history therefore acts as an internal representation (Friston, 2005, 2010; Teufel and Fletcher, 2020) that stabilizes perception when sensory signals are ambiguous. Intriguingly, we found that the balance between perceptual history and disambiguating sensory information slowly alternates between internally and externally oriented modes of sensory processing. Computational modeling indicated that between-mode transitions were likely to be caused by fluctuations in how strongly perception was driven by the accumulating effects of perceptual history.

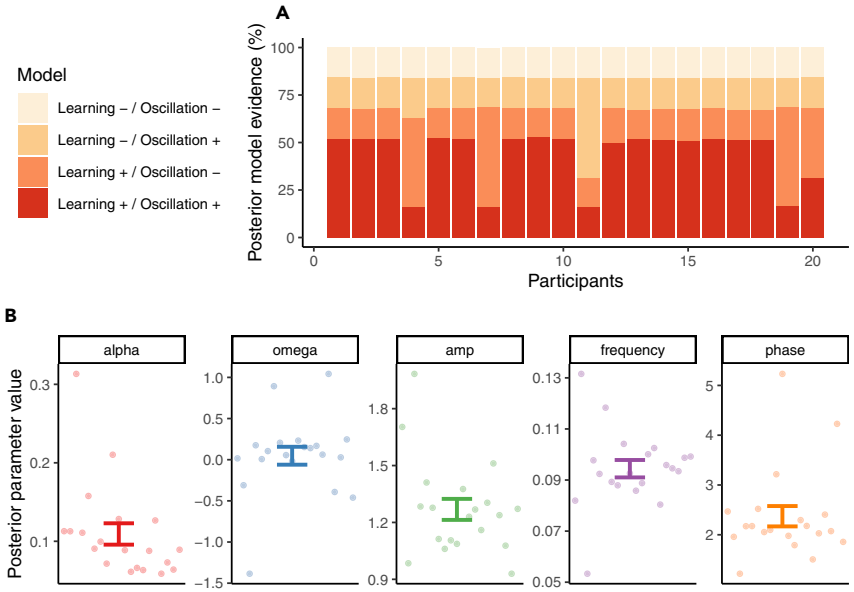


Figure 4. Computational modeling: Results

(A) **Model-level inference.** Random-effects Bayesian model selection identified $M_{Learning+/Oscillation+}$ as the clear winning model (group-level protected exceedance probability = 99.82%).

(B) **Parameter-level inference.** This model assumes that the external sensory signals is detected with a sensitivity parameter of $\alpha = 0.11 \pm 0.01$. The internal representation derived from perceptual history is updated as a function of the sequence of percepts according to learning rate $\omega = 0.05 \pm 0.11$. κ, the impact of accumulating perceptual history on perception, fluctuated according to a sine function with an amplitude of 1.27 ± 0.06, a frequency of $0.09 \pm 3.41 \times 10^{-3}$ (in *nb trials*$^{-1}$), and a phase of 2.37 ± 0.2. Pooled data are represented as mean $\pm$ SEM.

It may be argued that temporally extended biases toward internal information are not generic but specific to the class of structure-from-motion stimuli (Longuet-Higgins, 1986) investigated here. Indeed, structure-from-motion induces relatively long perceptual dominance durations (Weilnhammer et al, 2014, 2016, 2020). In addition, individual observers have been shown to exhibit stable idiosyncratic biases toward one of the two stimulus interpretations (Mamassian and Wallace, 2010; Weilnhammer et al., 2020), which can become strong enough to override disambiguating 3D cues (Wexler et al., 2015).

In this work, however, two factors speak against the view that transitions to internal mode were caused exclusively by strong perceptual biases. Firstly, we found relatively weak imbalances between the two possible states induced by our partially ambiguous structure-from-motion stimulus (see Results section). Secondly, we observed frequent transitions from internal to external mode while sensory information was held constant at threshold (see Figure 2), arguing against stable biases as the primary determinant of internally biased processing during graded ambiguity. Yet, to empirically assess this caveat, future work should investigate whether between-mode transitions occur also for ambiguous stimuli that induce shorter dominance durations, such as the Necker cube (Kornmeier and Bach, 2005). This would help understand whether fluctuations between internal and external mode depend on the type, strength, and temporal characteristics of bistable perception or, alternatively, occur independently of these factors and thus constitute a more general feature of perceptual processing.

As a second alternative explanation of our results, it may be proposed that fluctuating biases toward internal or external mode do not represent a perceptual phenomenon but, conversely, occur only due to processes that are situated downstream of perception, such as changes in reporting behavior (Brascamp et al., 2018) that are caused by periodic changes in how well participants attended to the experimental task (Zalta et al., 2020). Our analysis of response times (Figure S4), which are classically linked to fluctuating attention in paradigms such as the *Continuous Performance Task* (Rosenberg et al., 2013), did not yield any evidence for systematic differences in response behavior between internal and external mode. Yet, future experiments should apply no-report paradigms (Frässle et al., 2014), pupillometry (Lawson et al., 2020), or experimental manipulations of on-task attention (Alais et al., 2010) to dissociate post-perceptual processes from the perceptual phenomenon of mode-switching proposed in this work.

In a similar vein, it may be argued that slow fluctuations between externally and internally biased perception reflect epiphenomena that may arise from arbitrary constraints of neural processing (Honey et al., 2017). On the other hand, there may also be a specific computational benefit to slow transitions between external and internal model (Honey et al., 2017; Palva and Palva, 2011; VanRullen, 2016): In stable environments, internal mode may come with the benefit of a dramatic reduction in the energy demands of perception (Friston, 2010). Periodic switches to external mode may ensure that internal representations are updated in response to potential changes in the environment (Honey et al., 2017). In contrast to simultaneous processing, periodic mode switches may allow the brain to differentiate between internal and external sources of information (Honey et al., 2017). This may help perception to solve the credit-assignment problem, i.e., deciding whether to update internal representations of the environment or, alternatively, to modify beliefs about the reliability of sensory information (Weilnhammer et al., 2018). Thus, mode switching may represent a process that helps constructing stable representations of the environment despite ongoing sensory inputs (Bengio et al., 2015).

Indeed, fluctuations between externally and internally biased processing have been described in a variety of cognitive domains, including perception (Monto et al., 2008), episodic memory (Duncan et al., 2012), and waking state (McGinley et al., 2015). Switching between external and internal processing modes may thus represent a general computational mechanism that helps to adaptively integrate prior predictions with new information (Honey et al., 2017). Alterations in the temporal dynamics of mode switching may therefore represent the neurocomputational basis of psychotic experiences that often co-occur across cognitive domains, such as hallucinations, delusions, and altered sense of agency (Horga and Abi-Dargham, 2019; Sterzer et al., 2018).

To test the hypothesis that mode switches represent an adaptive mechanism that occurs across cognitive domains, future research should investigate whether transitions between external and internal mode can be induced experimentally. Based on the results of our computational modeling analysis, it may be hypothesized that participants should be more prone to transition from internal to external mode when repeatedly confronted with information that contradicts past experiences. Conversely, transitions from external to internal mode should occur more swiftly when participants receive information that is in line with prior predictions. Further down the line, it may be speculated that the overall frequency of mode switches could be altered by experimentally manipulating the volatility of the input data (Iglesias et al., 2013; Mathys et al., 2014).

Likewise, the existence of mode switches should be further substantiated by investigating whether external and internal modes can be determined based on markers that are independent of the perceptual response such as pupillary response or heart rate (Lawson et al., 2020). Together with an experimental manipulation of between-mode transitions, such markers could help to understand whether transitions between internal and external modes indeed represent an adaptive cognitive strategy that aids learning, or, alternatively, result from independent phenomena such as adaption (Chopin and Mamassian, 2012), attention (Alais et al., 2010), or response behavior (Frässle et al., 2014).

Limitations of the study

In this study, we have shown that bistable perception cycles through prolonged periods of enhanced and reduced sensitivity to disambiguating stimulus information. This finding suggests that conscious experience is characterized by slow fluctuations between internally and externally oriented modes of sensory processing. As a first limitation, our work investigated between-mode transitions only for a specific class of bistable stimuli (ambiguous structure-from-motion). Future work should test whether alternations between internally and externally oriented modes of processing also occur in other bistable stimuli, in particular in relation to paradigms that induce shorter dominance durations. As a second limitation, our work defines internal and external modes solely on the basis of behavior. Future studies should apply independent markers for internal and external modes (such as pupillometry) to probe between-mode transitions irrespective of behavioral reports. As a third limitation, our study does not provide an experimental control of pre- and post-perceptual processes such as attention or response behavior. Future experiments should use no-report paradigms or experimental manipulations of on-task attention to confirm that mode-switching represents a perceptual phenomenon rather than a process that occurs up- or downstream of perception.

Resource availability

Lead contact

The lead contact for this study is Veith Weilnhammer (veith-andreas.weilnhammer@charite.de).

Material availability

Materials have been deposited at OSF: https://doi.org/10.17605/OSF.IO/Y2CFM.

Data and code availability

Original data and code have been deposited at OSF: https://doi.org/10.17605/OSF.IO/Y2CFM.

METHODS

All methods can be found in the accompanying transparent methods supplemental file.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2021.102234.

ACKNOWLEDGMENTS

Author VW is a fellow of the Clinician Scientist Program funded by the Charité – Universitätsmedizin Berlin and the Berlin Institute of Health. This program was initiated and led by Prof. Dr. Duska Dragun to enable physicians to pursue a parallel career in academic research. With great sadness we have received the news that Prof. Dragun passed away on December 28th of 2020. We dedicate this publication to her as a mentor, friend, role model, and stellar scientist.

PS is funded by the German Research Foundation (STE 1430/8-1) and the German Ministry for Research and Education (ERA-NET NEURON program 01EW2007A). We acknowledge support from the German Research Foundation (DFG) and the Open Access Publication Fund of Charité – Universitätsmedizin Berlin.

AUTHOR CONTRIBUTIONS

- VW and PS conceptualized the study.
- VW designed the experiment.
- VW and MC collected the data.
- VW and PS wrote the initial draft and edited the manuscript.
- VW, MC, and PS reviewed the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: October 28, 2020
Revised: January 20, 2021
Accepted: February 21, 2021
Published: March 19, 2021

SUPPORTING CITATIONS

The following references appear in the supplemental information: Gekas et al., 2019.

REFERENCES

Abrahamyan, A., Silva, L.L., Dakin, S.C., Carandini, M., and Gardner, J.L. (2016). Adaptable history biases in human perceptual decisions. Proc. Natl. Acad. Sci. U S A *113*, E3548–E3557.

Alais, D., Boxtel, J.J.van, Parker, A., and Ee, R.van (2010). Attending to auditory signals slows visual alternations in binocular rivalry. Vis. Res. *50*, 929–935.

Bengio, Y., Lee, D.-H., Bornschein, J., Mesnard, T., and Lin, Z. (2015). Towards biologically plausible deep learning. arRxiv, 1502.04156.

Bergen, R.S.van, and Jehee, J.F. (2019). Probabilistic representation in human visual cortex reflects uncertainty in serial decisions. J. Neurosci. *39*, 8164–8176.

Brascamp, J., Sterzer, P., Blake, R., and Knapen, T. (2018). Multistable perception and the role of the frontoparietal cortex in perceptual inference. Annu. Rev. Psychol. *69*, 77–103.

Brascamp, J.W., Knapen, T.H.J., Kanai, R., Noest, A.J., van Ee, R., and van den Berg, A.V. (2008). Multi-timescale perceptual history resolves visual ambiguity. PLoS One *3*, e1497.

Chopin, A., and Mamassian, P. (2012). Predictive properties of visual adaptation. Curr. Biol. *22*, 622–626.

Duncan, K., Sadanand, A., and Davachi, L. (2012). Memory's Penumbra: episodic memory decisions induce lingering mnemonic biases. Science *337*, 485–487.

Fischer, J., and Whitney, D. (2014). Serial dependence in visual perception. Nat. Neurosci. *17*, 738–743.

Frässle, S., Sommer, J., Jansen, A., Naber, M., and Einhäuser, W. (2014). Binocular rivalry:

frontal activity relates to introspection and action but not to perception. J. Neurosci. *34*, 1738–1747.

Friston, K. (2010). The free-energy principle: a unified brain theory? Nat. Rev. Neurosci. *11*, 127–138.

Friston, K. (2005). A theory of cortical responses. Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci. *360*, 815–836.

Fritsche, M., Mostert, P., and Lange, F.P.de (2017). Opposite effects of recent history on perception and decision. Curr. Biol. *27*, 590–595.

Fründ, I., Wichmann, F.A., and Macke, J.H. (2014). Quantifying the effect of intertrial dependence on perceptual decisions. J. Vis. *14*, 1–16.

Gekas, N., McDermott, K.C., and Mamassian, P. (2019). Disambiguating serial effects of multiple timescales. J. Vis. *19*, 1–14.

Hanley, J.A. (2003). Statistical analysis of correlated data using generalized estimating equations: an orientation. Am. J. Epidemiol. *157*, 364–375.

Honey, C.J., Newman, E.L., and Schapiro, A.C. (2017). Switching between internal and external modes: a multiscale learning principle. Netw. Neurosci. *1*, 339–356.

Horga, G., and Abi-Dargham, A. (2019). An integrative framework for perceptual disturbances in psychosis. Nat. Rev. Neurosci. *20*, 763–778.

Hsu, S.M., and Wu, Z.R. (2020). The roles of preceding stimuli and preceding responses on assimilative and contrastive sequential effects during facial expression perception. Cogn. Emot. *34*, 890–905.

Iglesias, S., Mathys, C., Brodersen, K.H., Kasper, L., Piccirelli, M., Ouden, H.E.den, and Stephan, K.E. (2013). Hierarchical prediction errors in midbrain and basal forebrain during sensory learning. Neuron *80*, 519–530.

Kornmeier, J., and Bach, M. (2005). The Necker cube–an ambiguous figure disambiguated in early visual processing. Vis. Res. *45*, 955–960.

Körding, K.P., and Wolpert, D.M. (2004). Bayesian integration in sensorimotor learning. Nature *427*, 244–247.

Lawson, R.P., Bisby, J., Nord, C.L., Burgess, N., and Rees, G. (2020). The computational, pharmacological, and physiological determinants of sensory learning under uncertainty. Curr. Biol. *31*, 163–172.e4.

Leopold, D.A., Wilke, M., Maier, A., and Logothetis, N.K. (2002). Stable perception of visually ambiguous patterns. Nat. Neurosci. *5*, 605–609.

Liberman, A., Fischer, J., and Whitney, D. (2014). Serial dependence in the perception of faces. Curr. Biol. *24*, 2569–2574.

Logothetis, N.K., Leopold, D.A., and Sheinberg, D.L. (1996). What is rivalling during binocular rivalry? Nature *380*, 621–624.

Longuet-Higgins, H.C. (1986). Visual motion ambiguity. Vis. Res. *26*, 181–183.

Maloney, L.T., Dal Martello, M.F., Sahm, C., and Spillmann, L. (2005). Past trials influence perception of ambiguous motion quartets through pattern completion. Proc. Natl. Acad. Sci. U S A *102*, 3164–3169.

Mamassian, P., and Wallace, J.M. (2010). Sustained directional biases in motion transparency. J. Vis. *10*, 23.

Mathys, C.D., Lomakina, E.I., Daunizeau, J., Iglesias, S., Brodersen, K.H., Friston, K.J., and Stephan, K.E. (2014). Uncertainty in perception and the hierarchical Gaussian filter. Front. Hum. Neurosci. *8*, 825.

McGinley, M.J., Vinck, M., Reimer, J., Batista-Brito, R., Zagha, E., Cadwell, C.R., Tolias, A.S., Cardin, J.A., and McCormick, D.A. (2015). Waking state: rapid variations modulate neural and behavioral responses. Neuron *87*, 1143–1161.

Monto, S., Palva, S., Voipio, J., and Palva, J.M. (2008). Very slow EEG fluctuations predict the dynamics of stimulus detection and oscillation amplitudes in humans. J. Neurosci. *28*, 8268–8272.

Palva, J.M., and Palva, S. (2011). Roles of multiscale brain activity fluctuations in shaping the variability and dynamics of psychophysical performance. In Progress in Brain Research (Elsevier B.V.), pp. 335–350.

Pearson, J., and Brascamp, J. (2008). Sensory memory for ambiguous vision. Trends Cogn. Sci. (Regul. Ed.) *12*, 334–341.

Powers, A.R., Mathys, C., and Corlett, P.R. (2017). Pavlovian conditioning–induced hallucinations result from overweighting of perceptual priors. Science *357*, 596–600.

Prado, J., Carp, J., and Weissman, D.H. (2011). Variations of response time in a selective attention task are linked to variations of functional connectivity in the attentional network. NeuroImage *54*, 541–549.

Rosenberg, M., Noonan, S., DeGutis, J., and Esterman, M. (2013). Sustaining visual attention in the face of distraction: a novel gradual-onset continuous performance task. Atten. Percept. Psycho. *75*, 426–439.

Stephan, K.E., Penny, W.D., Daunizeau, J., Moran, R.J., and Friston, K.J. (2009). Bayesian model selection for group studies. NeuroImage *46*, 1004–1017.

Sterzer, P., Adams, R.A., Fletcher, P., Frith, C., Lawrie, S.M., Muckli, L., Petrovic, P., Uhlhaas, P., Voss, M., and Corlett, P.R. (2018). The predictive coding account of psychosis. Biol. Psychiatr. *84*, 634–643.

Teufel, C., and Fletcher, P.C. (2020). Forms of prediction in the nervous system. Nat. Rev. Neurosci. *21*, 231–242.

Urai, A.E., Braun, A., and Donner, T.H. (2017). Pupil-linked arousal is driven by decision uncertainty and alters serial choice bias. Nat. Commun. *8*, 14637.

Urai, A.E., De Gee, J.W., Tsetsos, K., and Donner, T.H. (2019). Choice history biases subsequent evidence accumulation. eLife *8*, e46331.

VanRullen, R. (2016). Perceptual cycles. Trends Cogn. Sci. *20*, 723–735.

Weilnhammer, V., Röd, L., Eckert, A.-L., Stuke, H., Heinz, A., and Sterzer, P. (2020). Psychotic experiences in schizophrenia and sensitivity to sensory evidence. Schizophrenia Bull. *46*, 927–936.

Weilnhammer, V.A., Ludwig, K., Sterzer, P., and Hesselmann, G. (2014). Revisiting the Lissajous figure as a tool to study bistable perception. Vis. Res. *98*, 107–112.

Weilnhammer, V.A., Sterzer, P., and Hesselmann, G. (2016). Perceptual stability of the lissajous figure is modulated by the speed of illusory rotation. PLoS one *11*, e0160772.

Weilnhammer, V.A., Stuke, H., Sterzer, P., and Schmack, K. (2018). The neural correlates of hierarchical predictions for perceptual decisions. J. Neurosci. *38*, 5008–5021.

Wexler, M., Duyck, M., and Mamassian, P. (2015). Persistent states in vision break universality and time invariance. Proc. Natl. Acad. Sci. U S A *112*, 14990–14995.

Wilson, R.C., and Collins, A.G. (2019). Ten simple rules for the computational modeling of behavioral data. eLife *8*, e49547.

Zalta, A., Petkoski, S., and Morillon, B. (2020). Natural rhythms of periodic temporal attention. Nat. Commun. *11*, 1–12.

iScience, Volume *24*

Supplemental information

Bistable perception alternates between internal and external modes of sensory processing

Veith Weilnhammer, Meera Chikermane, and Philipp Sterzer

Supplementary Information

"Blasts from the past: Bistable perception alternates between internal and external modes of sensory processing."

Authors:

Veith Weilnhammer[1,2], Meera Chikermane[1], Philipp Sterzer[1,2,3,4]

Affiliations and Contribution:

[1] Department of Psychiatry, Charité-Universitätsmedizin Berlin, corporate member of Freie Universität Berlin and Humboldt-Universität zu Berlin, 10117 Berlin, Germany

[2] Berlin Institute of Health, Charité-Universitätsmedizin Berlin and Max Delbrück Center, 10178 Berlin, Germany

[3] Bernstein Center for Computational Neuroscience, Charité-Universitätsmedizin Berlin, 10117 Berlin, Germany

[4] Berlin School of Mind and Brain, Humboldt-Universität zu Berlin, 10099 Berlin, Germany

Corresponding Author:

Veith Weilnhammer, Department of Psychiatry, Charité Campus Mitte, Charitéplatz 1, 10117 Berlin, phone: 0049 (0)30 450 517 317, email: veith-andreas.weilnhammer@charite.de

Word Count: 3415 words (main text without methods)

OSF-Project: https://osf.io/y2cfm/

1 Transparent Methods

1.1 Participants

We recruited a total of 20 participants (9 female; age: 27.45 ± 1.01 years). All participants had (corrected-to-) normal vision, were naive to the purpose of the study and gave informed, written consent prior to the experiment authorized by the Charité ethics committee.

1.2 Apparatus

Stimuli were presented on a 98PDF-CRT-Monitor (60 Hz, 1040 x 1050 pixels, 60 cm viewing distance, 41.38 pixels per degree [°] visual angle) using Psychtoolbox 3 and Matlab R2007b (MathWorks). 3D stimulation was achieved using 3D red-blue filter glasses. The blue filter was placed over the right eye.

1.3 Heterochromatic Flicker Photometry

Subjective differences in luminance can induce 3D effects based on the Pulfrich effect. To preclude that this phenomenon induces biases with regard to direction of rotation in partially ambiguous structure-from-motion stimuli, we conducted a separate pre-test experiment. We used *Heterochromatic Flicker Photometry* to estimate subjective equiluminance between red and blue. We presented red and blue circles (diameter: 6.45° visual angle) alternating at a frequency of 15 Hz. In case of subjective differences in luminance, participants perceived a flicker, which they reduced by adjusting the luminance of the blue stimulus initially presented at a random luminance between 0 and 125% relative to the red stimulus presented at a fixed luminance of 100%. Average equiluminance estimated across 10 such trials determined the monitor- and participant-specific luminance of the red- and blue-channels (average blue

luminance: 110.85 ± 4.74%).

1.4 Main experiment

The main experiment assessed how perceptual history was integrated with varying levels of disambiguating sensory information. To induce bistable perception, we generated rotating discontinuous structure-from-motion stimuli by placing a total of 2000 dots (each subtending 0.08° visual angle, overall stimulus size: 14.5° x 14.5°) on the surface of a Lissajous band (see Figure 1A and additional Supplementary Video V1). The Lissajous band was formed by the perpendicular intersection of two sinusoids ($x(t) = sin(A * t)$ and $y(t) = cos(B * t + \delta)$ with $A = 3$, $B = 8$). Within each trial, the stimulus was presented for 2 sec, while δ increased from 0 to 0.5π. The width of the Lissajous band was set to 0.04π ° rotational angle. Fixation intervals between trials were uniformly jittered around 2.5 ± 0.25 sec.

To generate parametric 3D stimuli, we attached a stereo-disparity signal to a fraction of the dots on the Lissajous band. Dots that carried stereo-disparity information were represented on separate monocular channels. To this end, corresponding pairs of red (left eye) and blue (right eye) dots were shifted against each other by 0.01π rotational angle. Dots without stereo-disparity information were presented binocularly. The wavelength of binocular dots (purple) was defined by adding the individual wavelengths of red and blue (see Heterochromatic Flicker Photometry). Throughout the experiment, we varied the signal-to-ambiguity ratio (SAR) by manipulating the fraction of dots that carried stereo-disparity information (see below). The direction of disambiguating sensory information (i.e., whether the front surface of the partially disambiguated sphere moved to the left or to the right) was randomized across trials.

We instructed participants to indicate the perceived direction of rotation of the Lissajous band by pressing the arrow-keys on a standard keyboard (right index finger: rotation of the front surface of the Lissajous to the *left*; right ring finger: rotation to the *right*; middle finger: *unclear* or mixed direction of rotation). *Error* responses were defined for trials at which

participants did not respond before the end of stimulus presentation or indicated more than one perceptual response.

Within one run, participants viewed a total of 120 trials. In runs R1-4, we adjusted the SAR dynamically based on a staircase procedure (Figure 1A; see Gekas et al. for a similar approach that manipulated the ambiguity of Gabor stimuli by parametrically varying their orientation (Gekas et al., 2019)). To this end, we defined *checkpoint-trials* that occurred in intervals of 10 trials, starting at the 11th trial of each run. At each checkpoint-trial, we computed the number of stimulus-congruent trials in the block of 10 trials preceding the checkpoint-trial. If more than 8 trials within the preceding block were stimulus-congruent (i.e., perceived in congruence with disambiguating sensory information), we decreased the SAR for the upcoming block by 5%. For 8 stimulus-congruent trials, the SAR remained unchanged in the upcoming block. If we observed less than 8 and more than 5 stimulus-congruent trials, we increased the SAR by 5% in the upcoming block. For less than 6 stimulus-congruent trials, we increased the SAR by 10% in the upcoming block. Run R1 started at an initial SAR of 100%. Runs R2-4 started at the final SAR obtained in the preceding run. During the final runs R5 and R6, we fixed the SAR to the average SAR obtained during runs R1-4.

1.5 Analyses

As dependent variables-of-interest, we computed the proportion of stimulus-congruent trials (i.e., trials perceived in congruence with disambiguating sensory information) and history-congruent trials (i.e., trials perceived in congruence with the immediately preceding percept). At every trial, we recorded the specific perceptual response (left, right, unclear or error) and response time (difference between the button-press indicating the percept and trial onset). *Directed biases* in perception were assessed via the probability of trials perceived as rotating to the right (ranging from 0 to 100%). We computed *absolute biases* by taking the absolute difference between the probability of trials perceived as rotating to the right and chance level at 50%. For summary statistics, we computed the dependent variables within runs R1-6 or

for levels of SAR, respectively, and averaged across participants. For dynamic analyses, we computed the dependent variables at each trial within a sliding window of ± 5 trials. Trials were allocated to the *internal mode* of perceptual processing if the sliding probability of history-congruent percepts was above the sliding probability of stimulus-congruent percepts (vice versa for *external mode*). *Intermediate mode* was designed as a rest category accounting for the fraction of trials where the sliding probabilities of history- and stimulus-congruent percepts were equal (see Supplementary Figure 1C for a representative time course). In 6 participants, we detected runs in which no mode-switch occurred. These runs were excluded when computing the average duration between mode-switches.

Statistical procedures were carried in *R* (summary statistics) and *Matlab* (computational modeling). We conducted group-level pair-wise comparisons using two-sided paired t-tests. Differences from chance-level were evaluated using two-sided one-sample t-tests. We performed correlative analyses using Pearson correlation. We applied the R-method *glm* with a binomial link-function for logistic regression and used the R-packages *lmer* and *afex* for linear mixed effects modeling.

Bayes factors were computed using the R-package *BayesFactor*, using the function *ttestBF* and *lmBF* for linear models. For t-tests, we placed a noninformative Jeffreys prior on the variance of the normal population and a Cauchy prior (rscale = 0.71) on the standardized effect size. Linear models used g-priors (fixed effects: rscale = 0.71; random effects = 1). To obtain Bayes factors for main effects and interactions, we estimated full and reduced models and divided the respective Bayes Factors.

1.5.1 Logistic regression and simulation analyses

In simulation analyses, we asked whether logistic regression reproduced the overall probability of stimulus- and history-congruent percepts. Moreover, we used these simulations to test whether the Markovian assumption of logistic regression (i.e., that the percept at trial t depends exclusively on the current sensory information at trial t and the percept at the

immediately preceding trial *t-1*) could explain the observed fluctuations between external and internal modes of perceptual processing. Prior to simulation analysis, we estimated logistic regression models that predicted the perceptual response p at each trial t based the dependent variables h (perceptual history) and d (disambiguating sensory information):

$$p(t) = \beta_h * h(t) + \beta_d * d(t) \tag{1}$$

The dependent variable *perceptual history* (*h(t)*) was defined by the perceptual response *p(t)* (0: leftward rotation; 1: rightward rotation) at the preceding trial:

$$h(t) = p(t-1) \tag{2}$$

The dependent variable *Disambiguating sensory information* (*d(t)*) was defined by a linear transform of the *Direction of disambiguation* (*DIR*, 0: leftward rotation; 1: rightward rotation) and the signal-to-ambiguity ratio (*SAR*, ranging from 0 to 100%) at trial t:

$$d(t) = 0.5 + (DIR(t) - 0.5) * SAR/100 \tag{3}$$

By setting either β_h or β_d to zero, we created reduced logistic regression models that were compared based on Akaike Information Criterion (AIC). As indicated by equation (1), none of the logistic regression models contained an interaction term. For simulation, we used the full logistic regression model, with *SAR* set to the individual threshold SAR used in run R5 and 6. *DIR* was chosen at random for every simulated trial. In analogy to the actual experiment, we simulated 120 trials per run. The total number of simulated runs amounted to 1000 for each participant.

1.5.2 Computational modeling

We constructed all models using the Hierarchical Gaussian Filter toolbox (Mathys et al., 2014) as implemented in the HGF 4.0 toolbox (distributed within the TAPAS toolbox; https://www.tnu.ethz.ch/de/software/tapas). At each trial t, the possible perceptual states y were coded as

$$y(t) = \begin{cases} 1: & \rightarrow & (rotation) \\ 0: & \leftarrow & (rotation) \end{cases} \tag{4}$$

The input to the model u was provided a linear combination of the direction of disambiguation (DIR) and the signal-to-ambiguity ratio (SAR):

$$u(t) = 0.5 + (DIR(t) - 0.5) * SAR/100 \tag{5}$$

To predict the participants' trial-wise perceptual responses, we combined input u with the prior probability of the perceptual states $\hat{\mu_1}(t)$ into the first-level posterior μ_1.

$$\eta_1(t) = exp(-(u(t) - 1)^2/(2 * \alpha)) \tag{6}$$

$$\eta_0(t) = exp(-(u(t))^2/(2 * \alpha)) \tag{7}$$

$$\mu_1(t) = \frac{\hat{\mu_1}(t) * \eta_1(t)}{\hat{\mu_1}(t) * \eta_1(t) + (1 - \hat{\mu_1}(t)) * \eta_0(t)} \tag{8}$$

In these equations, the influence of disambiguating sensory information on perception scales with the sensitivity parameter α, which was estimated as a free parameter in all models. When α approaches zero, $\mu_1(t)$ is close to the binary values of $u(t)$ (i.e., 0: stimulation with

3D-information for leftward rotation; 1: rightward rotation), signaling high sensitivity to sensory information. Conversely, for α increasing toward infitiy, $\mu_1(t)$ is close to $\hat{\mu_1}(t)$ (see below), reflecting low sensitivity to sensory information.

The influence of perceptual history, in turn, is represented by $\hat{\mu_1}(t)$. The value of $\hat{\mu_1}(t)$ depends on the dynamic accumulation of history effects in μ_2 (i.e, the estimated prior probability of perceptual states represented at the second level of the HFG), which represents the tendency of the first level posterior towards $\mu_1(t) = 1$. For higher values of κ, the prior probability of perceptual states μ_2 has a stronger impact on $\hat{\mu_1}(t)$. The influence of perceptual history on the participants' experience therefore scales with κ:

$$\hat{\mu_1}(t) = s(\kappa * \mu_2(t-1)) \tag{9}$$

Importantly, the models considered in this manuscript differ with respect to the computation of μ_2 (Dimension 1) and κ (Dimension 2).

1.5.2.1 Dimension 1

For models $M_{Learning+/Oscillation-}$ and $M_{Learning+/Oscillation+}$, μ_2 is updated via precision-weighted prediction errors that are generated by the sequence of perceptual states:

$$\mu_2(t) = \hat{\mu_2}(t) + \frac{1}{\pi_2(t)} * \delta_1(t) \tag{10}$$

$$\hat{\mu_2}(t) = \mu_2(t-1) \tag{11}$$

The precision of the second-level representation of perceptual history is governed by $\pi_2(t)$

$$\pi_2(t) = \hat{\pi_2}(t) + \frac{1}{\hat{\pi_1}(t)} \tag{12}$$

The difference between the first level perceptual prediction $\hat{\mu_1}(t)$ and the first-level posterior $\mu_1(t)$ yields the prediction error $\delta_1(t)$:

$$\delta_1(t) = \mu(t) - \hat{\mu}_1(t) \tag{13}$$

$\delta_1(t)$ is combined with the second level precision π_2, yielding the precision-weighted prediction error $\epsilon_2(t)$, which updates the second level prediction $\hat{\mu}_2(t)$:

$$\epsilon_2(t) = \frac{1}{\pi_2} * \delta_1(t) \tag{14}$$

In addition to κ and α, $M_{Learning+/Oscillation+}$ and $M_{Learning+/Oscillation-}$ incorporate a learning rate ω. This free parameter determines how swiftly μ_2 is updated in response to predicition errors, thereby controlling the speed at which the second-level precision $\hat{\pi_2}(t)$ changes over time.

$$\hat{\pi_1}(t) = \frac{1}{\hat{\mu_1}(t) * (1 - \hat{\mu_1}(t))} \tag{15}$$

$$\hat{\pi_2}(t) = \frac{1}{\frac{1}{\pi_2(t)} + \exp(\omega_2)} \tag{16}$$

By contrast, for models $M_{Learning-/Oscillation-}$ and $M_{Learning-/Oscillating+}$, $\mu_2(t)$ is defined by the immediately preceding perceptual state:

$$\mu_2(t) = \begin{cases} 1: & y(t) = 1 \\ -1: & y(t) = 0 \end{cases} \tag{17}$$

Thus, $M_{Learning-/Oscillation-}$ and $M_{Learning-/Oscillating+}$ do not incorporate any second-level accumulation of perceptual history and are thus governed only by the parameters κ and α.

1.5.2.2 Dimension 2

For models $M_{Learning-/Oscillation-}$ and $M_{Learning+/Oscillating-}$, κ is estimated as a stable parameter. By contrast, for models $M_{Learning-/Oscillation+}$ and $M_{Learning+/Oscillating+}$, κ fluctuates dynamically according to the frequency parameter f (in $nb\ trials^{-1}$), the phase parameter p and the amplitude parameter amp

$$\kappa = \frac{(amp * sin(f * t + p) + 1)}{2} \tag{18}$$

1.5.3 Model inversion

We used a free energy minimization approach for model inversion (Friston, 2010), maximizing a lower bound on the log-model evidence for the individual participants' data. Parameters were optimized using quasi-Newton Broyden-Fletcher-Goldfarb-Shanno minimization. Parameters were inverted using the following priors:

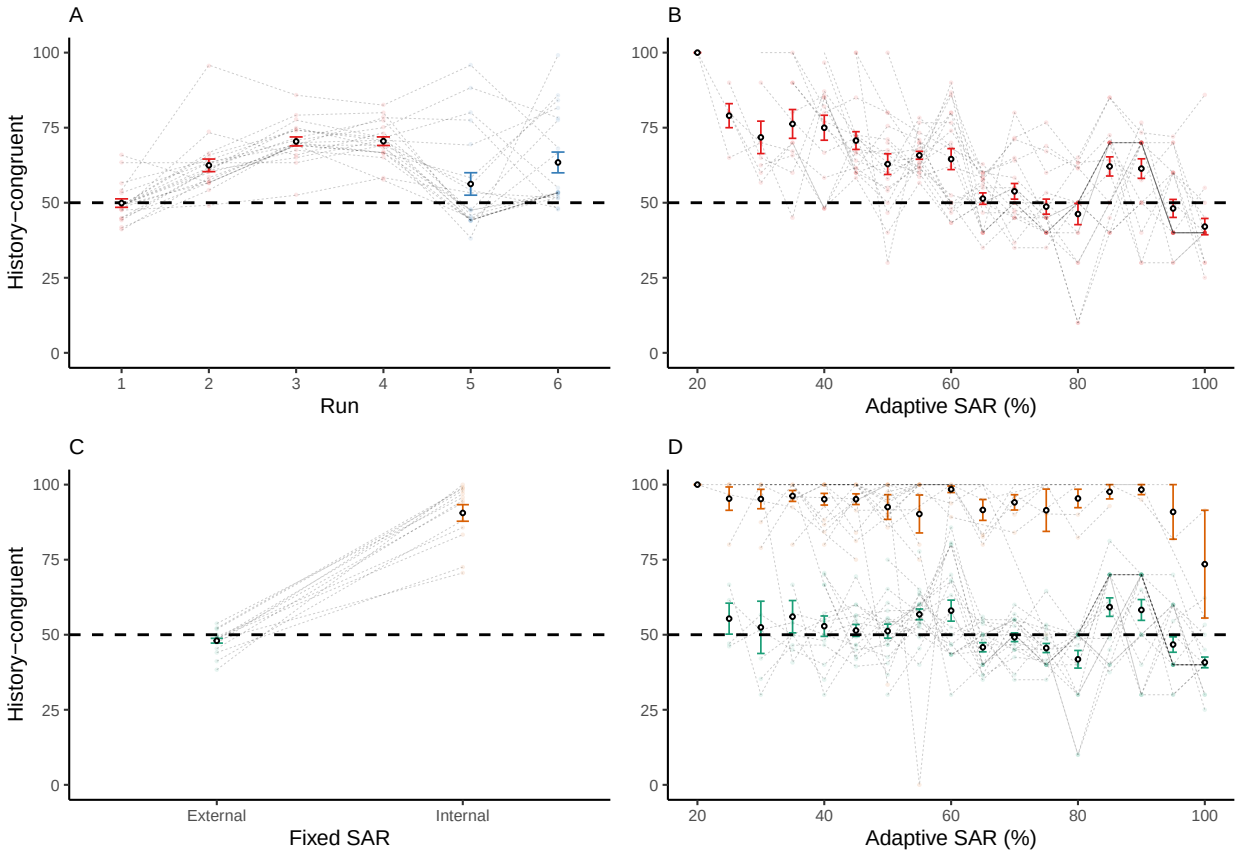
- Dimension 1: κ = prior mean of log(1) and prior variance of 1; α = prior mean of log(0.1) and prior variance of 1; ω = prior mean of 0 and prior variance of 16.
- Dimension 2: α = prior mean of log(0.1) and prior variance of 1; f = prior mean of log(0.1 and prior variance of 0.1; p = prior mean of $\pi/2$ and prior variance of $\pi/2$; amp = prior mean of log(1) and prior variance of 1.

1.5.4 Model-level inference

Models were compared using random-effects Bayesian model selection (Stephan et al., 2009) as implemented in SPM12 (http://www.fil.ion.ucl.ac.uk/spm/software/spm12). We report protected exceedance probabilities for group-level inference and individual exceedance probabilities at the participant-level.

2 Supplementary Figures

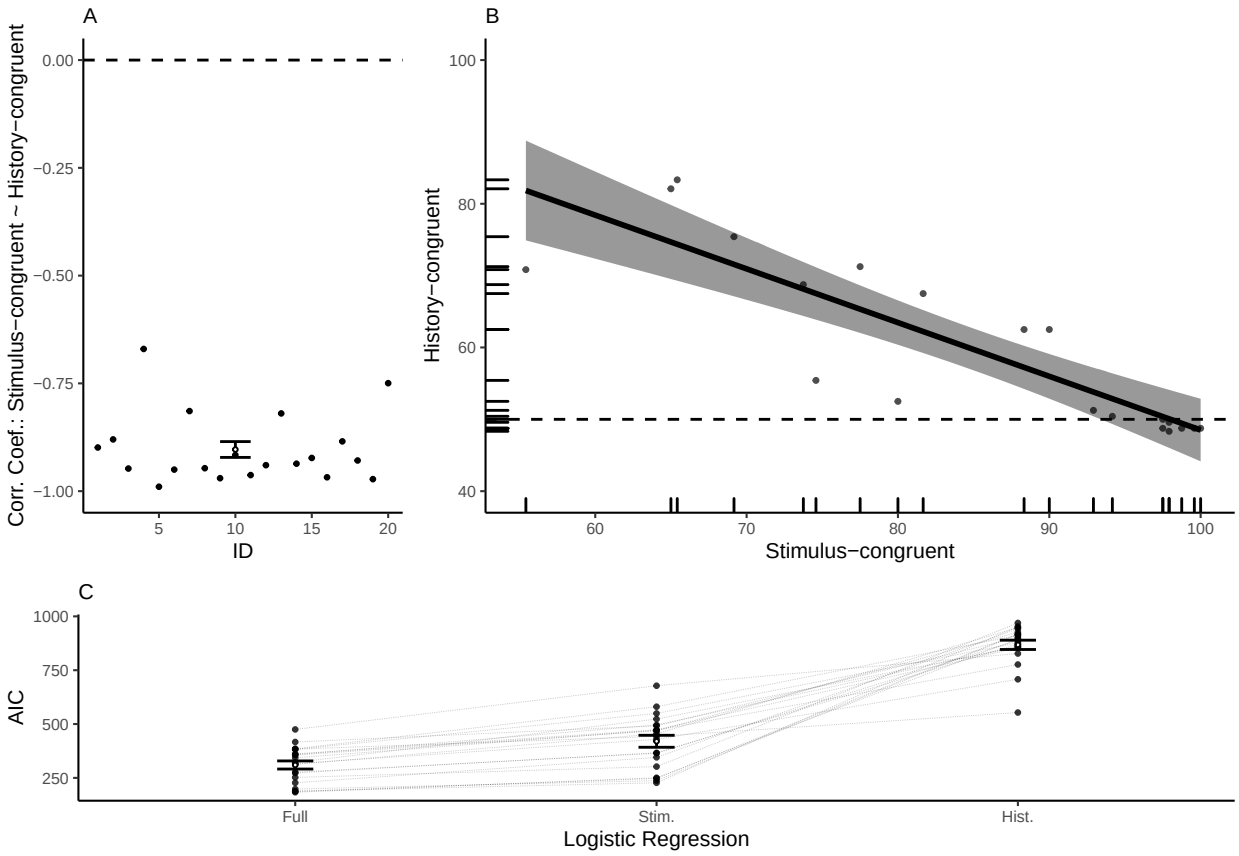
2.1 Supplementary Figure S1



Supplementary Figure S1. A. Perceptual history across runs. Related to Figure 1 and 2. As the SAR was dynamically adjusted in runs R1-4 (shown in red), we observed a progressive increase in the frequency of history-congruent percepts ($F(3, 57) = 57.96$, $p = 2.58 \times 10^{-17}$, $BF_{10} = 9.69 \times 10^{15}$; R1: 49.88 ± 1.41%; R2: 62.46 ± 2.1%; R3: 70.42 ± 1.5%; R4: 70.5 ± 1.43%). In runs with fixed SAR (R5-6, depicted in blue), history-congruent percepts amounted to 56.25 ± 3.74% in R5 and 63.42 ± 3.46% in R6. **B. Perceptual history across levels of SAR.** As expected, perceptual history had a stronger influence on perception at lower levels of SAR ($F(1, 265.07) = 181.5$, $p = 7.25 \times 10^{-32}$, $BF_{10} = 5.2 \times 10^{28}$, main effect of *SAR*) and ranged at chance-level when disambiguating sensory information was strong. **C. Perceptual history during internal and external mode for SARs at**

threshold. During internal mode, the frequency of history-congruent percepts approached 100% (90.57 ± 2.76%), but was reduced below chance level during external mode (48 ± 0.8%; T(19) = -2.51, p = 0.02, BF_{10} = 2.74, one-sample t-test). **D. History-congruent percepts during internal and external mode across the full range of SAR.** Linear mixed effects modeling indicated that frequency of history-congruent percepts was significantly affected by the factor *mode* (green: external; yellow: internal; F(2, 484.03) = 23.87, p = 1.3×10^{-10}, $BF_{10} = 2.43 \times 10^{70}$) and showed a trend for an effect of *SAR* (F(1, 188.7) = 3.42, p = 0.07, BF_{10} = 1.1). We observed no between-factor interaction with respect to history-congruent percepts (F(2, 469.91) = 0.07, p = 0.93, BF_{10} = 0.05). Please note that any main effect of *mode* was expected, since external and internal mode were defined based on the dynamic probability of stimulus-congruence. Pooled data are represented as mean ± SEM.

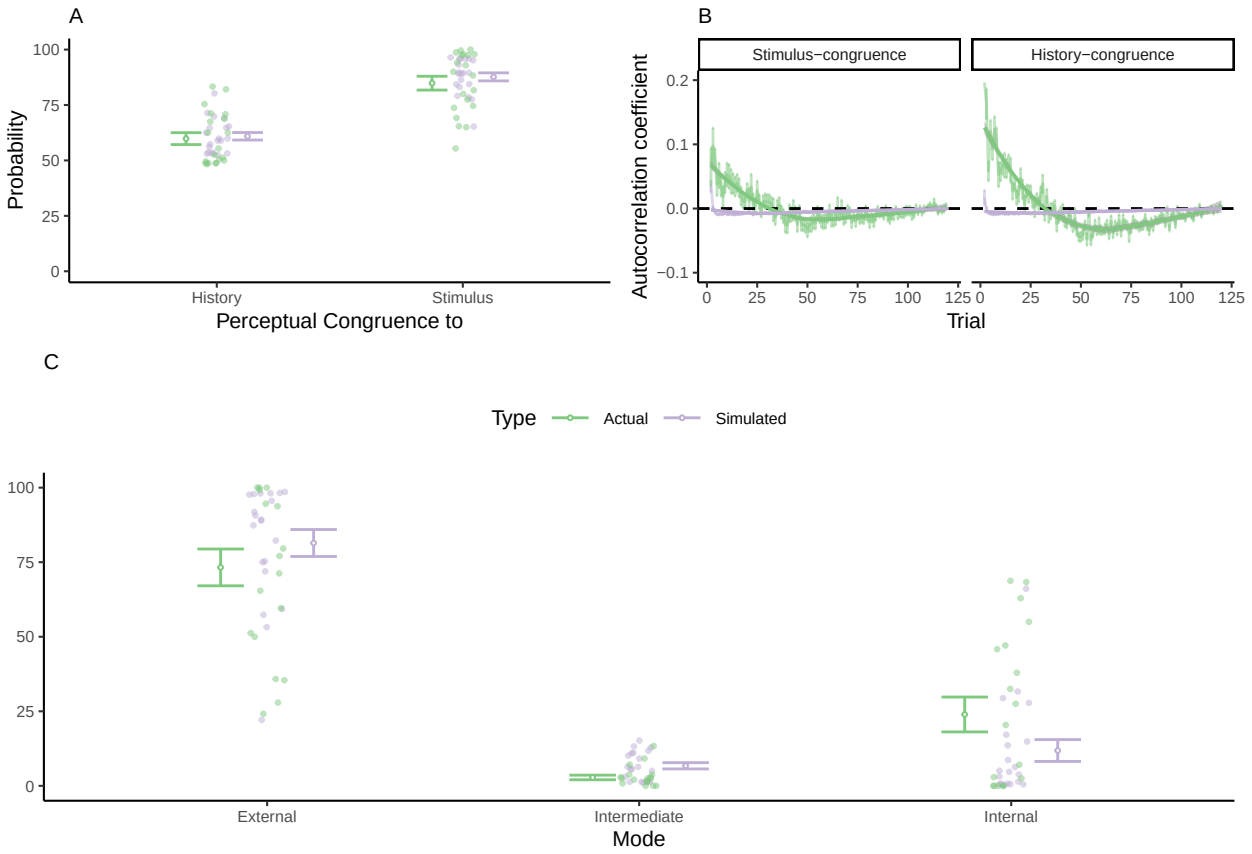
2.2 Supplementary Figure S2



Supplementary Figure S2. A. Within-participant correlations of stimulus- and history-congruent percepts. Related to Figure 1 and 2. In individual participants, Pearson correlation coefficients between the frequencies of stimulus- and history-congruent percepts (runs R5-6; fixed SAR) amounted to -0.9 ± 0.02 (T(19) = -49.25, p = 1.66×10^{-21}, $BF_{10} = 1.34 \times 10^{18}$, one-sample t-test). This strong inverse relationship suggested that perceptual history and disambiguating sensory information compete with each other to determine conscious experience. **B. Across-participants correlation of stimulus- and history-congruent percepts.** Inter-individual differences in the frequency of history-congruent percepts strongly predicted the frequency of stimulus-congruent percepts (ρ = -0.77, $p = 7.2 \times 10^{-5}$, $BF_{10} = 203.27$, Pearson correlation for runs R5-6). This negative association indicated that, overall, perceptual history had a stronger impact in participants who were less sensitive to disambiguating sensory information. **C. Predicting perceptual**

responses using logistic regression. In each participant, the Akaike Information Criterion (AIC) of logistic regression models based on both disambiguating sensory information and perceptual history (309.95 ± 18.81) was lower than the AIC for models based on sensory information only (419.95 ± 27.84; T(19) = -9.39, p = 1.45×10^{-8}, $BF_{10} = 8.89 \times 10^{5}$, paired t-test) or perceptual history only (867.86 ± 21.58; T(19) = -16.46, p = 1.06×10^{-12}, $BF_{10} = 6.54 \times 10^{9}$). Pooled data are represented as mean ± SEM.

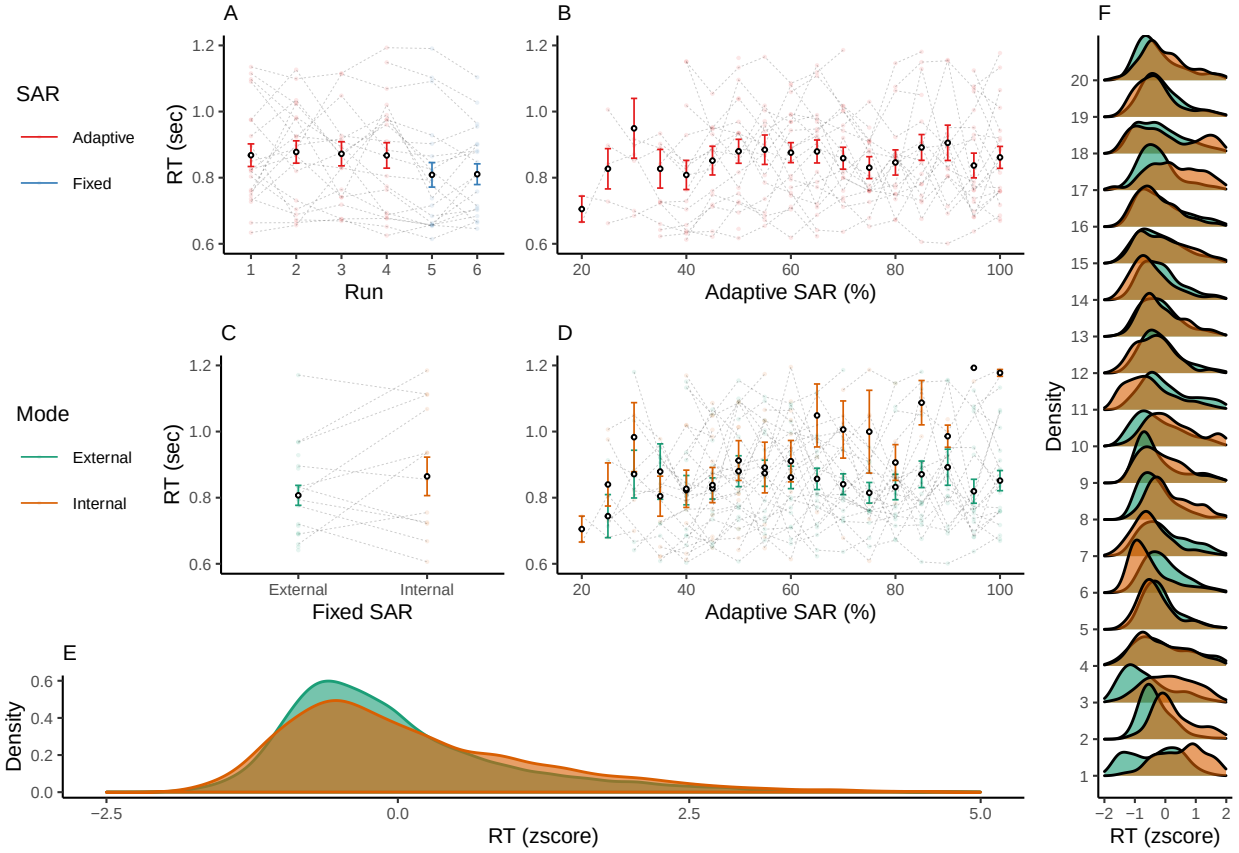
2.3 Supplementary Figure S3



Supplementary Figure S3. A. Simulating the overall frequencies of stimulus- and history-congruent percepts with logistic regression. Related to Figure 1 and 2. We estimated logistic regression models based on both disambiguating sensory information and perceptual history in individual participants and used the regression weights to simulate perceptual responses. These simulations revealed that logistic regression reproduced the overall frequencies of history-congruent percepts observed in the actual experiment (simulated data in purple: 60.9 ± 1.69%; actual data in light green: 59.83 ± 2.69%; T(19) = 0.78, p = 0.44, $BF_{10} = 0.31$, paired t-test) as well as the overall frequency of stimulus-congruent percepts (simulated: 87.69 ± 1.81%; actual data: 84.85 ± 3.12%; T(19) = 1.48, p = 0.16, $BF_{10} = 0.59$). **B. Simulated autocorrelations of stimulus- and history-congruence.** When simulating perceptual responses from logistic regression, we detected no autocorrelation of stimulus- or history-congruence. Real trial-wise autocorrelation coefficients are plotted

for comparison. **C. Simulating the relative proportions of external, internal and intermediate modes with logistic regression.** Likewise, logistic regression did not reproduce the relative proportion of trials spent in external mode (simulated: 81.43 ± 4.52%; actual: 73.25 ± 6.17%; T(19) = 2.75, p = 0.01, $BF_{10} = 4.17$, paired t-test), internal mode (simulated: 11.85 ± 3.66%; actual: 23.94 ± 5.84%; T(19) = -3.49, p = 2.44×10^{-3}, $BF_{10} = 16.92$) and intermediate mode (simulated: 6.72 ± 1.05%; actual: 2.81 ± 0.77%; T(19) = 3.73, p = 1.41×10^{-3}, $BF_{10} = 27.07$). Pooled data are represented as mean ± SEM.

2.4 Supplementary Figure S4



Supplementary Figure S4. A. RTs across runs. Related to Figure 1 and 2. In runs R1-4 (depicted in red), we adapted the SAR based on a staircase procedure, which did not affect RTs (R1: 0.87 ± 0.03 sec; R2: 0.88 ± 0.03 sec; R3: 0.87 ± 0.04 sec; R4: 0.87 ± 0.04 sec). In runs R5-6 (depicted in blue), the SAR was fixed to the average SAR from the preceding runs R1-4 (60.25 ± 2.36 sec). RTs amounted to 0.81 ± 0.04 sec in R5 and 0.81 ± 0.03 sec in R6. **B. RTs across levels of SAR.** Globally, the level of disambiguating sensory information did not have a significant effect on RT ($F(1, 261.5) = 0.05$, $p = 0.82$, $BF_{10} = 0.15$, main effect of SAR). **C. RTs during internal and external mode for SAR at threshold.** In Runs R5 and R6, RTs did not differ between external and internal mode ($T(12) = 0.74$, $p = 0.48$, $BF_{10} = 0.35$, paired t-test): **D. RTs during internal and external mode across the full range of SAR.** Linear mixed effects modeling indicated that, during internal mode, response times increased significantly for higher levels of SAR

(F(2, 476.5) = 10.73, p = 2.77×10^{-5}, $BF_{10} = 538.42$, *mode* x *SAR* interaction), driving a main effect of *SAR* in this analysis (F(1, 488.29) = 21.98, p = 3.57×10^{-6}, $BF_{10} = 1.73$). Response times were longer during internal mode (F(2, 474.05) = 5.28, p = 5.39×10^{-3}, $BF_{10} = 18.9$, main effect of *mode*). **E. Collapsed RTs.** During both internal and external mode, normalized RTs were better explained by a log-normal distribution (internal mode: AIC = 1.07×10^4, external mode: AIC = 2.79×10^4) as compared to a Gaussian distribution (internal mode: AIC = 1.08×10^4, external mode: AIC = 2.81×10^4), a gamma distribution (internal mode: AIC = 1.08×10^4, external mode: AIC = 2.8×10^4) or a Weibull distribution (internal mode: AIC = 1.24×10^4, external mode: AIC = 3.39×10^4). **F. Individiual RT distributions.** Within individual participants (y-axis), the distributions of normalized RTs were largely overlapping between internal and external mode. Pooled data are represented as mean ± SEM.