# Loss of precise auditory sampling as a sign of value-driven visual attentional capture

Rodrigo Caramés Harcevnicow[1,*] Thaiz Sánchez-Costa[1,*], Alejandra Carboni[1,2], and Francisco Cervantes Constantino[1,3,‡]

[1]Centro de Investigación Básica en Psicología; Universidad de la República; Montevideo, Uruguay
[2]Instituto de Fundamentos y Métodos en Psicología; Universidad de la República; Montevideo, Uruguay
[3]Instituto de Investigaciones Biológicas “Clemente Estable”; Montevideo, Uruguay

*These authors contributed equally
‡Corresponding address:
Centro de Investigación Básica en Psicología
Facultad de Psicología, Universidad de la República
Dr. Tristán Narvaja 1674
Montevideo 11200, Uruguay
fcervantes@psico.edu.uy

## Abstract

By linking with rewards, sensory signals gain salience and the ability to influence selective attention even when they have become irrelevant. The dynamics of reward-driven distraction in the brain remains unclear, particularly at the time of shaping multisensory associations. It is unknown whether reward-driven distraction by visual signals interferes with the robust ability of the brain to phase-lock to auditory modulations. In a dynamic audiovisual (AV) coherence task, we investigated how visual reward associations affect performance and modulate auditory encoding precision. Participants were presented with dual visual object streams flickering at different rates, accompanied by an amplitude-modulated sound matching one of the flicker periods, for subjects to identify the matching visual object. At the periphery, an irrelevant color feature flickers in sync with the target and may capture observers’ attention, due to a prior color-reward association training regime. Electroencephalography (EEG) recordings assessed participants' sensitivity to the audiovisual task. The findings indicate that target discrimination was impoverished in the presence of colors that had previously been associated with reward. The phase locking of auditory responses also decreased, evidencing an attentional shift away from auditory modulation representations. Moreover, down-modulations of auditory phase locking predicted the effect size of participants' reward-driven distraction. These findings highlight how value-driven attentional capture fundamentally alters multimodal processing in the temporal domain. They suggest that less precise neural representations of unisensory streams not connected to reward-associated cues undermine the processing of temporal coherence relationships between multisensory streams. Momentary inter-modal competition, induced by reward-driven distraction, appears consistent with the systematic exploit of gaps in active attentional sampling strategies that unfold over time.

**Keywords: value-driven attentional capture – reward-driven distraction – auditory encoding – phase locking value**

## 1. Introduction

Goodies and tidbits make tangible souvenirs of past experiences. Even after our priorities change, signals associated with rewards can distract us, a phenomenon known as value-driven attentional capture (VDAC) or reward-driven distraction (Anderson, 2019; Anderson et al., 2011, 2021). How we tune to rewarded components in the environment is shaped by the context and history of reward associations that we may have experienced in the past (Todd & Manaligod, 2018). When a stimulus comprises certain features that were previously associated with reward, such as a color or sound, this may elicit distraction even when those features are not currently rewarding, and furthermore, it may adjust the share of neural processing relative to current targets (Rusz et al., 2020). Moreover, such influences can be transferred from signals in one sensory modality to another. In one example, reward-associated but irrelevant tone distractors shape the visuospatial encoding of co-localized images; the altered balance of visual processing may predict VDAC behavior (Pooresmaeili et al., 2014). Thus, reward associations established within one sensory modality can impact both encoding and perceptual decisions in another when multisensory events are spatially linked.

It remains unclear whether VDAC similarly transfers its influence cross-modally when multimodal correspondences are dynamic and operate on temporal principles. Temporal associations play a crucial role in binding multisensory streams into stable multimodal objects in naturalistic settings (Bizley et al., 2016). Moreover, individuals' attentional sampling of their sensory environment follows precise spatial and temporal rules (Chen & Gong, 2022). Here, we aim to understand whether VDAC impacts the precision of sensory processing when multisensory correspondences are based on temporally coherent auditory and visual streams. In particular, the present study addresses whether precisely phase-locked auditory processing of sound modulations can be subject to cross-modal influences by the temporal dynamics of concurrent ‘has-been’ visual input.

Large population activity in the auditory cortex (AC) typically entrains to the slower temporal modulations of sound (e.g., < 12 Hz) with high temporal precision, but is this ability potentially resilient against cross-modal value-driven attentional capture (VDAC) from visual stimuli? Auditory representations can be strongly shaped by the dynamics of temporally coherent visual signals, for instance, by guiding segmentation of the auditory stream towards multimodally consistent principles (Atilgan et al., 2018; Maddox et al., 2015). The AC receives cross-sensory influences through supramodal systems such as the temporo-parieto-occipital junction (Rolls et al., 2022), of which its superior temporal aspect is involved in the auditory reward-driven capture of visual signals (Pooresmaeili et al., 2014). This suggests that precise temporal encoding in the auditory cortical systems may be altered under VDAC. We examine the phase information content of sensory-driven neural oscillations, a parameter used to index the strength of AV temporal relationships perceptually (Kösem et al., 2014). Phase studies have shown that attending to a unimodal input stream may adjust the ongoing oscillatory contexts of the primary cortices of other modalities (Lakatos et al., 2009) and that such phase resetting may generate apparent saliency changes (Kayser, 2009). Accordingly, we hypothesize that the phase locking of auditory networks to a sound pattern may be modulated, under VDAC, by salient visual signals.

In the present study, subjects were instructed to match the steady oscillations of an amplitude-modulated (AM) sound stream to one of two central visual flicker options, with the only correct choice beating at the same AM rate of sound. To induce value-driven attentional capture (VDAC), we simultaneously presented colored dot distractors at the visual periphery, which flickered alongside the central figures. Behaviorally, we predicted that trials presented with a high-reward color would capture more attention away from the target figures compared to colors lacking a reward association. At the neural level, we investigated how the phase consistency of neural oscillations, in response to auditory input, would be maintained in and out of the presence of these reward-associated colors. We presented the dynamics of central and peripheral patterns identically to prevent the potential confound that auditory activity is differentially entrained simply because of mismatching dynamics between central and peripheral patterns. We tested the hypothesis that visual colors that signal past reward associations may positively or negatively impact the consistent processing of auditory temporal information.

To generate reward associations, prior to this main AV test, we trained participants on a standard response inhibition task (Eriksen & Eriksen, 1974). At this *training* stage, trials were presented with peripheral colors that predicted task earnings, with a temporary reward structure that controlled for different color-reward mappings. The main AV task was subsequently used as a *testing* stage to investigate subject performance using reaction time and sensitivity measurements. Using frequency tagging methods with electroencephalography (EEG) recordings (Grahek et al., 2021), we probed changes to the modulations of temporal consistency (phase coherence) of neural activity in the auditory system in the presence of the different reward-associated visual signals in the AV task. As reward associations are typically learned in probabilistic designs (Anderson et al., 2011; Hickey & Peelen, 2015; Pooresmaeili et al., 2014; Sanz et al., 2018), but see (Anderson, 2016; Tankelevitch et al., 2020), we additionally evaluated whether the effects may be linked to reward uncertainty during learning (Cho & Cho, 2021; Koenig et al., 2017). For instance, reward associations that are induced with greater levels of variability may be more easily recognized at a later stage (Rouhani et al., 2018). One possible explanation is that such a greater range of variance modulates attention and thus the quality of sensory representations during training (Monosov, 2020).

In brief, our results demonstrate that value-driven cross-modal capture of visual attention interferes with auditory neural processing by hindering its precise synchronization to temporal modulations of sound. As participants progressed in the AV task, their performance became relatively impoverished in the presence of reward-associated distractors. Diminished phase locking of responses to the sound stream was associated with visual distractors signaling higher reward levels during training. Regardless of variability, these neural changes predicted behavioral differences in the AV task sensitivity at the testing stage. The findings also underscore the variable precision of auditory representations across the temporal domain and its impact on successfully framing the correspondences that determine AV integration.

## 2. Materials and methods

**2.1 Participants.** Thirty-four subjects (15 female; mean age 26.4 ± 4.0 SD; 4 left-handed) with no history of neurological or psychiatric disorders voluntarily participated in the study. All participants provided formal written informed consent. They reported normal hearing and normal or corrected to normal visual acuity. The experiments were performed in accordance with the WMA Declaration of Helsinki guidelines . The Ethics in Research Committee of the School of Psychology at the Universidad de la República approved the experimental procedures. EEG data from one participant did not contain sound output feedback recordings (see *Experimental equipment and setup*, below); for two other subjects, a half of the EEG sessions' recordings were not saved in error, reducing analyses to 31 subjects.

**2.2 Experimental equipment and setup.** Participants were seated at 50 cm in front of a 40 cm CRT monitor (E. Systems, Inc., CA), on which visual presentations were delivered with an 83 dpi resolution and an 85 Hz refresh rate. Audio presentations were delivered with a Sound Blaster Z sound card (Creative Labs, Singapore) in combination with a Scarlett 4i4 hub interface (Focusrite Plc, UK) and high-quality Sennheiser HD 25 headphones (Sennheiser, Germany). Neural recordings were performed using a BioSemi ActiveTwo 64-channel EEG system (BioSemi, The Netherlands) with a 10/20 CMS/DRL (ground) layout at a 2048 Hz digitization rate. Parallel audio signal sampling was obtained from sound interface output fed back to the ActiveTwo system through an optical link cable (ERGO Optical, BioSemi). A fifth-order cascaded integrator-comb low-pass filter with a −3 dB point at 410 Hz was applied to sensor channels online, after which all signals were decimated to 1024 Hz. Online high-pass response was fully DC-coupled. Participants adjusted the sound volume presentation to a comfortable listening level prior to starting the computerized tasks. Full experimental sessions lasted ~2.5 h. At the time of data collection, the following guidelines were implemented: the experimenter and participants wore a face mask at all times; the experimenters wore a face shield and disposable hospital gown during personal interaction with participants within the lab room; and electrode gel cap preparation was undertaken at an open-air facility. The participant performed the task with the lab room doors left open.

**2.3 Experimental protocol.** The main experimental session consisted of an initial Training task (Tr1) followed by an initial AV Testing task (Tt1). Thereafter, participants performed a second round of Training (Tr2) and Testing (Tt2) tasks (Figure 1A). Presentation and response logging were performed with PsychoPy (Peirce, 2007) for the Training tasks and with MATLAB R2010a (Natick, United States) for the Testing tasks.

**2.4 Training tasks (Tr1 & Tr2).** *2.4.1 Stimuli.* Participants responded to an arrow-based version of the Eriksen flanker task  that requires them to quickly and accurately report the direction of a central target (Figure 1B). 216 trials were presented in each Training task, with balanced congruent, incongruent, and neutral flanker conditions. A static colored ring display was shown at the start of a trial. The ring’s internal-external diameter was 21–27 degrees of visual angle (dva) and displayed in either solid red (‘r’, hexadecimal #FF0000), green (‘g’, #00FF00), blue (‘b’, #0000FF), cyan (‘c’, #00FFFF), magenta (‘m’, #FF00FF), or yellow (‘y’, #FFFF00) over a black background, allocating 36 trials per color. After an initial 200 ms of ring presentation, the white target and flanker symbols appeared, subtending 6 dva horizontally. Both the symbols and colored ring were shown for a further 940 ms, after which the ring disappeared. The symbols remained for an additional 50 ms, and there was a final 1.5 s response timeout period. Participants responded to the direction of the central symbol by pressing a keyboard button. Missed or late responses were treated as invalid trials.

*2.4.2 Reward structure.* Points were awarded for valid and correct trials, and awarded amounts were determined by ring color. Three reward levels (High, Low, No Reward) and two uncertainty levels (uncertainty/no, uncertainty/yes) determined earnings as follows: A flat 80-point reward was given for valid responses in High reward and no uncertainty conditions (‘High-un’); 20 points in Low reward and no uncertainty condition (‘Low-un’) trials. In High reward and uncertainty conditions (‘High-uy’), valid responses earned a number of points that was pseudo-randomly sampled from a normal distribution (mean = 80 points; standard deviation = 7 points) and rounded to the nearest integer. Low reward and uncertainty (‘Low-uy’) trials were similarly awarded, with the difference that the distribution mean was 20 points. ‘No Reward’ conditions always yielded 0 points and were used as baseline. To balance for color-specific effects, colors were paired with conditions per subject following a four-group cohort schedule (Table 1). Incorrect trials earned 0 points, and invalid trials were deducted 50 points. Points accumulated over the training task. Tr1 and Tr2 were identical in their reward structure. Trial conditions and colors were presented in random order. An ad lib pause was made every 72 trials, indicating the number of accumulated points so far. The order of trials was randomized for each task.

| Group cohort | High-uy | Low-uy | High-un | Low-un | No Reward |
|---|---|---|---|---|---|
| 1 | r | g | c | m | b,y |
| 2 | g | r | m | c | b,y |
| 3 | c | m | r | g | b,y |
| 4 | m | c | g | r | b,y |

**Table 1. Color-reward associations.** Participants received points on correct trials based on a high/low reward and a binary yes/no uncertainty scheme that was determined by the color of a ring displayed during the trial. To balance color-reward associations, each participant was assigned to one of four different group cohorts. For instance, in cohort 3, participants always received 80 points on the correct trials if a red ring was shown (‘High-un’ condition). Participants in cohort 4 instead earn the same reward at green ring trials. For blue and yellow trials, participants never earned points.

*2.4.3 Feedback.* After trial response or extinction, participants received text feedback with the proviso that, for Tr1, it appeared scrambled and illegible; feedback was clearly shown in Tr2 (Figure 1B). In consequence, the reward structure of the task was not available for participants to learn at the Tr1 stage but was at Tr2. Pause-related feedback and feedback from invalid trials that resulted in point deductions were always clearly presented. In both tasks, participants were instructed to accumulate as many points as possible. They were informed that the points would be converted into tokens at the end of each Testing task, to be exchanged for a prize at the end of the session.

**2.5 Testing tasks (Tt1/Tt2).** As part of a two-alternative forced choice (2AFC) audiovisual task, participants were presented with a visual flicker display that was simultaneously combined with an auditory amplitude-modulated (AM) noise stimulus. *2.5.1 Visual flicker display*. On each trial, two symbols were taken at random from a standard chess pool, i.e., king, queen, bishop, knight, rook, or pawn figures. Two replicates of each figure were shown (Figure 1C). Same-symbol pairs flickered at either 9 or 11 Hz, respectively. The flicker display was matched to the screen at an 85 Hz refresh rate via frequency approximation (Nakanishi et al., 2014). The four figure locations were scattered over a circular surface extending 8 dva in diameter and shown with a steady central fixation cross (Figure 1C). In addition, a peripheral ring-shaped aperture was displayed with similar dimensions as in the Training task. This domain was filled with a mixture of grey and colored dots. The specific color varied per trial and was from the same pool as in *Training*. The ratio of the gray set to the colored dot set was fixed at 4:3 (320 grey dots), and the diameter of each dot was 0.53 dva with a density of 2.6 dots/dva$^2$. To balance the general flicker between peripheral and central sets, both peripheral dot sets flickered, with the colored set exclusively synced to the target chess figure set and the gray set synced to the other chess figure (Video 1, Supplementary Figure 1). Dots and symbols were never overlain. The chess figures stayed in one location for 1 s before changing to different locations within the delimited central area. This resulted in a highly challenging task (see *Results*), and then for half of the participants, we presented the chess figures remaining at the same positions throughout the trial. The trial's display lasted 6 seconds in total.

*2.5.2 Auditory stimulus*. Auditory stimuli were constructed with MATLAB. Two 6 s AM white noise stimuli were synthesized with 9 and 11 Hz sinusoidal modulation rates, respectively, at a 75% modulation depth and sampled at 44.1 KHz. The same two auditory stimuli were used across trials and participants.

*2.5.3 Audiovisual presentation.* Each AV presentation consisted of the visual flicker display described above plus either the 9 or 11 Hz AM noise stimulus. Importantly, the AM rate also matched the flicker rate of the colored dots in every trial. In addition, the presentation's audio-video phase lag was randomized across trials. More specifically, the onset of the noise relative to the first video frame included a uniformly distributed delay in the (0,1/*f*) s range, where *f* is the sound AM rate, and was sampled every trial.

*2.5.4 AV test and feedback*. Participants were asked to accurately detect in each trial the chess symbol that matched the presented auditory AM pattern. Participants were explicitly instructed to ignore the colors and encouraged to maintain central fixation. After the AV display, the two symbols presented in the trial were shown again on the left and right sides of the screen. A click sound prompted the observer to respond by pressing 'a' or 'k' on the keyboard for the left or right side symbol, respectively. A high- or low-pitched feedback tone then indicated whether the response was correct or incorrect. Responses made before the click were not accepted, and accuracy was explicitly emphasized over speed for participants in their responses. Symbol pairs, colors, and auditory AM rates were equally distributed over 120 trials. Testing tasks Tt1 and Tt2 were identical. The order of trials was randomized for each task. At the end of the task, the percent correct performance was calculated and shown as feedback. This factor was multiplied by the number of points awarded at the immediately preceding Training task. This resulted in the number of tokens that subjects earned for the stage; for instance, 2000 points earned from Tr1 could be turned into 2 tokens at a 100% correct rate at Tt1. Participants were informed of the number of tokens accumulated so far at the end of each Testing task. As an incentive to improve their performance, at the start of Tt2, subjects were offered the possibility to pick among a set of award options if their next performance were to exceed that of Tt1. Accordingly, participants either selected or were assigned a chocolate bar among different types at the end of the session.

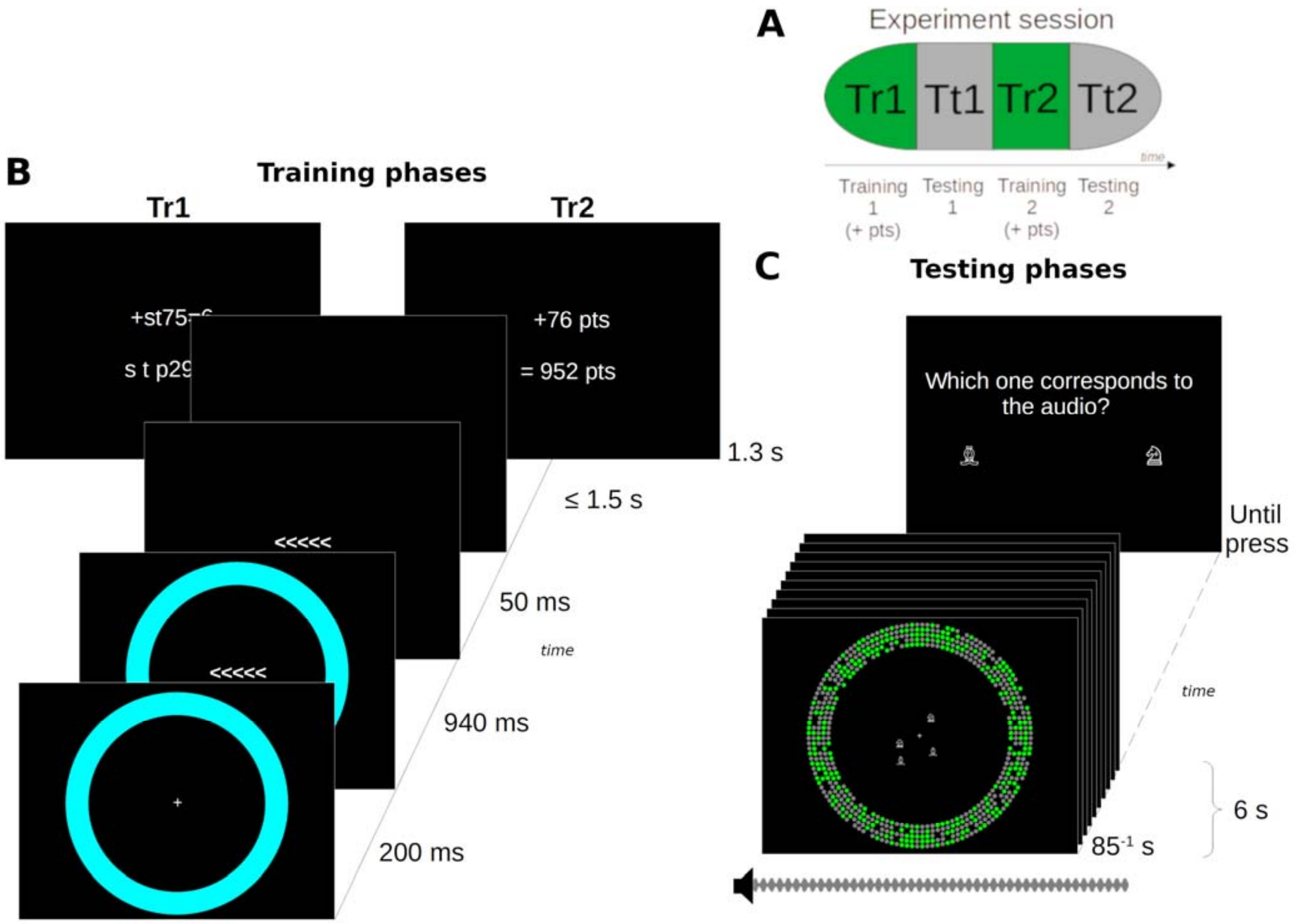

A
Experiment session
Tr1
Tt1
Tr2
Tt2
time
Training 1 (+ pts)
Testing 1
Training 2 (+ pts)
Testing 2
B
Training phases
Tr1
Tr2
+76 pts
= 952 pts
1.3 s
≤ 1.5 s
<<<<<
<<<<<
50 ms
time
940 ms
200 ms
C
Testing phases
Which one corresponds to the audio?
Until press
time
6 s
85⁻¹ s

**Figure 1. Value-driven attentional capture protocol.** (A) The full experiment consists of a first *Training* stage, then AV *Testing*, followed by a second round of Training and Testing. (B) At the Training stages, participants perform an Eriksen flanker task to accumulate points over the correct trials. The display includes a peripheral ring whose color predicts the number of points to be earned. Subject responses prompt feedback, but at the first training stage (Tr1), this information appears scrambled; at Tt2, feedback is clearly shown. (C) Participants are tested at two separate times (Tt1 and Tt2) in an identical 2AFC AV task containing a dynamic 6-second AV display. The display consists of one AM noise (shown at the bottom) plus two different visual symbols flickering at 9 and 11 Hz, respectively. The goal is to determine which of two chess objects corresponds to the same sound modulation rate, either 9 or 11 Hz. A peripheral ringed dot pattern is included, along with flickering grey and colored dots that are always in sync with corresponding chess symbols. At Tt2, colored dots signal reward associations learned from Tr2. See Video 1 for a full trial example.

**2.6 Data preprocessing.** Data analysis was implemented in MATLAB 2018b and concerns the Tt1 and Tt2 datasets. EEG recordings were downsampled to 512 Hz. Head and reference sensor datasets were common average-referenced, after which DC offset was removed per channel. Signals were bandpass-filtered between 1 and 40 Hz with a 10-order elliptic IIR filter of 60 dB attenuation and 0.5 dB passband ripple. Single channel data were rejected in a blind manner according to a channel variance-based criterion (Junghöfer et al., 2000) with confidence coefficient $\lambda_P = 4$ , and the same procedure was performed separately for the external reference channels. To reduce artifact contamination, EEG data were decomposed into independent components using FastICA (Hyvarinen, 1999). The two independent components with maximal broadband power were automatically selected and projected out of the data. A time-shifted principal component analysis (de Cheveigné & Simon, 2007) was applied to discard environmental signals recorded on the external reference sensors (±4 ms shifts). A sensor noise suppression algorithm (de Cheveigné & Simon, 2008) was applied in order to attenuate artifacts specific to any single channel (63 neighbors). After epoching, the variance-based rejection procedure was again applied across single channels and epochs' timeseries per participant, resulting in less than 2.9% rejected channel-epoch timeseries on average (subject range 2.7%–3.2%).

**2.7 Data analysis.** *2.7.1 Behavioral data*. D-prime (*d'*) scores indexing sensitivity to the audiovisual discrimination task were obtained from responses at Tt1 and Tt2 blocks. For each participant, *d'* measures were computed as the *z*-transform of the hit rates, minus the *z*-transform of false positives, to 9 Hz auditory stimulus presentations. In other words, correct 9 Hz trials contributed to overall hit rate scores, whilst incorrect 11 Hz trials were counted as false alarms. *z*-scores were obtained by the normal cumulative distribution inverse function approximation, and *d'* values were computed per condition, per subject. To ensure convergence, all hit rates and false alarm scores were approximated by the log-linear approach (Stanislaw & Todorov, 1999).

*2.7.2 EEG data*. In Tt1 and Tt2 blocks, single-trial EEG channel data were epoched according to the 6 s AV presentations and sub-epoched into 1 s partitions, then transformed into the frequency domain via a Fast Fourier Transform, which resulted in six spectra with 1 Hz resolution per trial. We used phase coherence analysis (Nozaradan et al., 2012; van Diepen & Mazaheri, 2018) to measure temporal and spectral phase synchronization in EEG signals across each trial's sub-epochs. Phase locking or synchronization values (PLV) are an effective means to test steady state responses in the brain as they are modulated by attention for stimulus-driven periodic paradigms (Keitel et al., 2019). Single-trial phase information was extracted from each sub-epoch spectrum, and the phase coherence spectrum of a trial was computed over the six sub-epochs via circular averaging of the corresponding phase angle timeseries. At a given frequency, coherence reaches a maximum value of 1 for perfectly phase-aligned sub-epochs and tends to 0 as the distribution of phases becomes uniform. To obtain auditory phase locking values (PLVs), peak sizes in the single-trial coherence spectrum were measured at the target auditory stimulation rate plus the first and second harmonics (i.e., 9, 18, and 27 Hz in 9 Hz trials; alternatively, 11, 22, and 33 Hz). Harmonic information was included since it contains non-redundant neural representations that complement those directly linked to the base rate (Jenkins et al., 2011; Keitel & Müller, 2016). Each peak size was expressed as the ratio of the PLV at the target frequency bin divided over the mean PLV at the two neighboring frequency bins. The auditory PLV for a trial was defined as the average of the three sound-related peak frequency ratios across nine auditory-dominant EEG channels (Fz, FC1, FCz, FC2, C3, C1, Cz, C2, C4, see Figure 2C) and expressed in decibels.

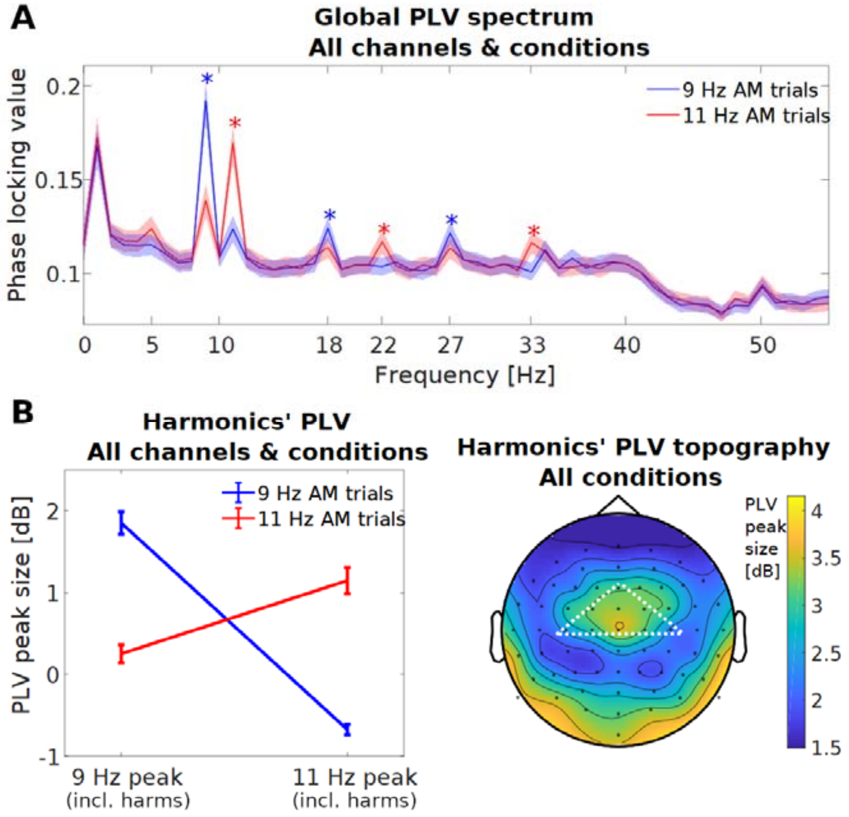


**Figure 2. Measuring the single-trial precision of auditory neural representations during AV scenes via AM phase locking.** (A) Phase locking values averaged across all conditions and channels. AV presentations involving 9 versus 11 Hz AM sound elicit characteristic frequency patterns in participants' global EEG responses, with significant PLV differences at the AM rate and two harmonics (adjusted *p*-values ≤ 0.004). (B) The precision of auditory neural representations is measured by the size of PLV peaks, estimated by referencing the crest value to the immediate neighboring bin frequencies. Expressed in dB, the average of these peak-to-baseline ratio estimates is highly specific to auditory stimulation (left). Error bars indicate ±1 SEM. Topographically, peak sizes show auditory processing over the frontocentral scalp of the EEG array (right). The highlighted area indicates channels that were used to index auditory precision in all subsequent analyses.

**2.8 Statistical analyses.** All statistical analyses were performed with MATLAB (R2018a) (The MathWorks, Natick MA), both with built-in functions as well as custom-written code. Effects were statistically assessed for confirmatory testing using linear mixed-effect (LME) models with the MATLAB function '`fitlme`'. The omnibus LME analyses evaluated the effects of independent variables "Stage" (Testing task 1 and Testing task 2), "Reward" level (High, average +80 points; Low, +20 points; and No Reward, +0 points), and "Uncertainty" (yes, variability of 7 standard deviations; and no, 0) on the dependent variable y (I: subject's d', II: neural PLV). The full model includes 8 fixed effects (intercept, Stage, Reward, Uncertainty, Stage:Reward, Stage:Uncertainty, Reward:Uncertainty, Stage:Reward:Uncertainty). To account for repeated measures within participants, the model includes random effects (intercept and slope) for each participant in the between-subjects design that involves two task difficulty levels. In Wilkinson notation, the full general model is expressed as: *y* □~□ Stage□ × □Reward □×□ Uncertainty□ +□ (Stage□×□Reward□×□Uncertainty | participant/difficulty). Satterthwaite approximations were used to calculate *p*-values and perform *F*-tests for each factor and their interactions in the mixed model. Where a significant effect of 'Stage' or interaction between fixed effects 'Stage' and 'Reward' and/or 'Uncertainty' was found, post hoc tests were carried out on data restricted to single reward levels or stages, with a Bonferroni-Holm correction to account for multiple comparisons.

In addition, exploratory analyses were performed to address whether observed neural effects are correlates of behavior in the task. To rule out the possibility that such correlations are driven by potential outliers, testing included robust correlation interval estimation (Pernet et al., 2013), Shepherd's correlation (Schwarzkopf et al., 2012), and robust regression with the Huber weighting function (Pooresmaeili et al., 2014). Our correlation findings (see *Results*) motivated an additional exploratory study to establish whether auditory PLV levels during VDAC, or else generally, bear a predictive relationship with outcomes at the single trial level. In this exploratory LME, we modeled the effects of the independent variable "Outcome" (Correct and Incorrect) on the dependent neural PLV variable. The exploratory model is expressed as: PLV □~□ Stage□ × □Reward × Outcome□ +□ (Stage□×□Reward□×□Outcome | participant/difficulty).

## 3. Results

We tested participants on a two-alternative forced choice (2AFC) AV coherence task (Figure 1C) with the aim of determining whether it is influenced by reward associations that are established through a conditioning training procedure (Figure 1B). The AV task was tested before and after reward conditioning. Our main findings concern the influence of reward association by irrelevant visual cues on AV performance and auditory processing.

*3.1 Behavioral results.* We evaluated whether subjects displayed any influence from the reward association procedure. The behavioral measures of audiovisual discrimination performance (*d*-prime) at the pre- and post-conditioning stages of AV testing were compared by means of a full linear mixed effects model. In this initial model, we hypothesized that such conditioning may be contingent on uncertainty in the reward training schedule (i.e., earning rewards on a flat versus stochastic schedule; see *Methods*). The analysis of *d'* response sensitivity measures by the factors 'Stage' (Tt1 versus Tt2), 'Reward' (High versus Low versus No Reward), 'Uncertainty' (Fixed versus Variable Reward), and their interactions, first confirmed that AV sensitivity changed between the two testing sessions. Participants improved their performance from Tt1 to Tt2 ($F_{1,33}$=7.50; $p$ = 0.010; Figure 3A), consistent with practice effects. However, the analysis also shows that this main effect on *d'* by 'Stage' was qualified by an interaction between 'Stage' and 'Reward' ($F_{1,35}$=4.54; $p$ = 0.040). Factor 'Uncertainty' and all other interactions were not significant (Table 2). The results thus suggest implicit reward learning in participants. To assess this, the significant interaction was analyzed in post hoc tests, which restricted AV performance by the Reward level associated with the peripheral dots, i.e., High, Low, and No Reward. The post hoc tests revealed significant performance improvement by factor 'Stage' only at the No Reward level ($F_{1,31}$=9.94; adj. $p$ = 0.011; Fig. 3A), as participants did not significantly increase their AV sensitivity at Tt2 under the two rewarded conditions (Table 3). The results hence indicate that benefits on performance gained from greater experience at the second testing round may be offset or not available in the presence of rewarded-associated signals, consistent with reward-driven distraction.

To facilitate comparison with studies in the literature of reward-driven distraction, we additionally computed the standardized mean change effect size (Rusz et al., 2020) of *d'* sensitivity measures at Tt2, contrasted between the High reward (mean *d'* = 0.80) and No Reward (mean *d'* = 1.05) conditions. The estimated effect size of 0.192 indicates that the presence of a high-reward color hindered task performance gains by almost a fifth of a standard deviation compared with no reward color distractors, a somewhat moderate effect compared to most visual studies but in line with cross-modal reports (Rusz et al., 2020).

### Omnibus F-test with Satterthwaite approximation analysis of behavioral performance

Model: *d'* ~ Stage*Reward*Uncertainty + (Stage*Reward*Uncertainty | participant\difficulty)

| Term | F-value | *p*-value |
|---|---|---|
| Intercept | | |
| Stage | $F_{1,33}=7.50$ | **0.010** |
| Reward | $F_{1,32}=1.23$ | 0.28 |
| Uncertainty | $F_{1,36}=0.09$ | 0.77 |
| Stage × Reward | $F_{1,35}=4.54$ | **0.040** |
| Stage × Uncertainty | $F_{1,34}=1.80$ | 0.19 |
| Reward × Uncertainty | $F_{1,31}=0.10$ | 0.76 |
| Stage × Reward × Uncertainty | $F_{1,32}=1.39$ | 0.25 |

Values in bold indicate *p*-values below 0.05 significance threshold

**Table 2.** Omnibus analysis of behavioral *d*-prime data.

### Post-hoc analyses of training stage on behavioral performance, per reward cue level

F-tests with Satterthwaite approximation test results

Model: *d'ᵢ* ~ Stage + ( Stage | participant\difficulty), where *i* = {+80, +20, +0} reward levels

| | +80 (High reward) | | +20 (Low reward) | | +0 (No reward) | |
|---|---|---|---|---|---|---|
| | F-value | Adj. *p*-value | F-value | Adj. *p*-value | F-value | Adj. *p*-value |
| Intercept | | | | | | |
| Stage | $F_{1,31}=0.29$ | >0.9 | $F_{1,31}=0.42$ | >0.9 | $F_{1,31}=9.94$ | **0.011** |

P-values corrected for multiple comparisons through Bonferroni-Holm correction

**Table 3.** Post hoc analyses of behavioral *d*-prime data.

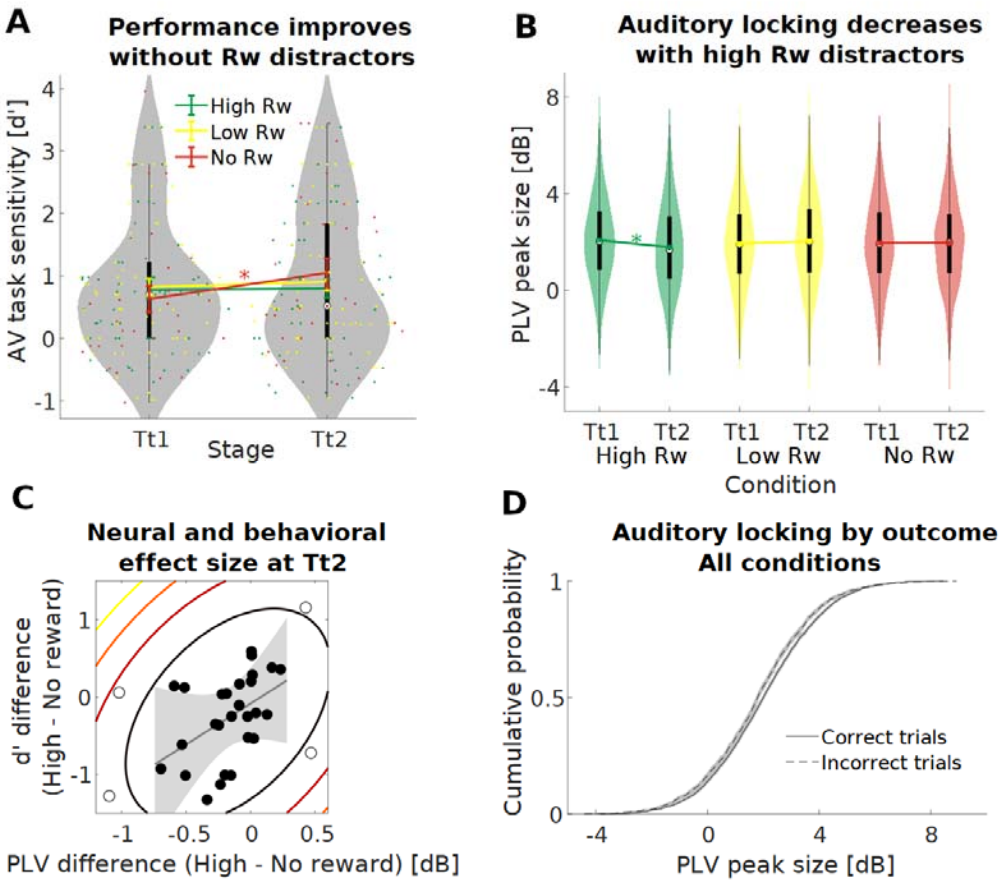


**Figure 3**. **Reward-driven visual distraction interferes with the behaviorally relevant precision of stimulus-locked auditory activity.** (A) Subjects perform the AV coherence task better with increased experience at Tt2, contingent on trial presentations being devoid of reward-associated color distractors. (B) Post-conditioning, AV task trials that present high-reward-associated colors result in lower neural phase-locking values to the auditory stimulation. Color distractors always flicker in phase with targets; hence, the retreat in precise auditory encoding is consistent with the multimodal segregation of sound representations given attended distractor information. Sub-optimal reward levels did not similarly influence the precision of auditory responses. (C) The loss in auditory precision predicts subjects' weaker behavioral sensitivity (reward-driven distraction) in the AV task. (D)

Exploratory findings that auditory precision measured via neural phase locking is a relevant decision variable in the AV task; auditory PLVs are modulated by the outcome at the single trial level across all conditions.

*3.2 Neural results.* We next studied the influence of the reward association on the precision of auditory responses to the AV scenes. Scalp-wide EEG activity during the AV scenes was generally locked to stimulation parameters, as expected by the stimulation design (Figure 2). To assess the temporal precision of auditory responses, we estimated the peak-to-baseline ratios (peak sizes) of EEG channels' PLV spectra at frequencies determined by auditory/target modulation rates (9, 18, and 27 Hz; alternatively, 11, 22, and 33 Hz). These rate peak magnitudes were found to map to a topographic EEG distribution predominantly consistent with auditory activity (Nozaradan et al., 2012). To study VDAC effects on temporal precision of auditory processing, we considered averaged PLV peak size data only from frontocentral auditory channels in all subsequent analyses (Figure 2C). Neural measures of auditory phase-locked responses (PLV peak size) stages were then compared, on a trial-by-trial basis, by means of a full linear mixed effects model that was structured as in behavioral analyses. Analysis of PLV peak size estimates by the factors 'Stage', 'Reward', 'Uncertainty', and their interactions confirmed a significant interaction between 'Stage' and 'Reward' ($F_{1,245}$=4.30; $p$ = 0.039) in these neural data as well. No other significant interaction or main effect was found (Table 4).

As before, the significant interaction was analyzed in post hoc tests that restricted PLV peak sizes by the Reward level associated with peripheral dots. The post hoc tests revealed a significant decrease in PLV by 'Stage', restricted only to the High reward condition ($F_{1,31}$=13.63; adj. $p$ = 0.003), with no evidence of change at all of the other reward conditions (Table 5). The results suggest that, if visual distractors signal considerable reward associations to observers, the ability of the auditory system becomes suboptimal to respond in a coherent fashion to sound modulations over the time such cues are present.

**Omnibus F-test with Satterthwaite approximation analysis of single-trial auditory PLV**

Model: PLV ~ Stage*Reward*Uncertainty + ( Stage*Reward*Uncertainty | participant\difficulty)

| Term | F-value | *p*-value |
|---|---|---|
| Intercept | | |
| Stage | $F_{1,70}$=1.64 | 0.20 |
| Reward | $F_{1,39}$=0.68 | 0.42 |
| Uncertainty | $F_{1,51}$=0.18 | 0.67 |
| Stage × Reward | $F_{1,245}$=4.30 | **0.039** |
| Stage × Uncertainty | $F_{1,156}$=0.75 | 0.39 |
| Reward × Uncertainty | $F_{1,38}$=0.32 | 0.57 |
| Stage × Reward × Uncertainty | $F_{1,474}$=0.01 | 0.90 |

Values in bold indicate *p*-values below 0.05 significance threshold

**Table 4**. Omnibus analysis of single-trial PLV peak size data.

**Post-hoc analyses of training stage on single-trial auditory PLV, per reward cue level**

F-tests with Satterthwaite approximation test results

Model: $\text{PLV}_i$ ~ Stage + ( Stage | participant\difficulty), where $i$ = {+80, +20, +0} reward levels

| | +80 (High reward) | | +20 (Low reward) | | +0 (No reward) | |
|---|---|---|---|---|---|---|
| | F-value | Adj. *p*-value | F-value | Adj. *p*-value | F-value | Adj. *p*-value |
| Intercept | | | | | | |
| Stage | $F_{1,31}$=13.63 | **0.003** | $F_{1,31}$=1.58 | 0.44 | $F_{1,207}$<0.01 | >0.9 |

P-values corrected for multiple comparisons through Bonferroni-Holm correction

**Table 5.** Post hoc analyses of single-trial auditory PLV data.

To find whether the behavioral changes may be related to neural modulations under VDAC, we examined whether, at a subject level, the scale of changes to temporal encoding evidenced by PLV predicted the behavioral performance differences. An effect size (either in *d'* or PLV) was defined by the separation between the High and No Reward conditions, the latter serving as the baseline. Computed at the key post-conditioning Tt2 stage, participants' changes on auditory PLV estimates were significantly and positively correlated with their behavioral differences (Spearman's *r*=0.473 [0.113–0.735 C.I.]; *p*=0.007; Fig. 3C). The correlation was robust to potential outliers (Shepherd's Pi=0.529; *p*=0.010), and a robust regression analysis additionally confirmed its significance (*t*=2.48; d.f.=29; *p*=0.019). In other words, manifesting steeper decreases in auditory PLV predicted greater loss of sensitivity to the task during reward-driven distraction. By contrast, the same exploratory analysis set at the earlier Tt1 stage (pre-conditioning) failed to reveal a significant correlation between neural and behavioral effects (r=0.228; [-0.171–0.556 C.I.]; *p*=0.22) in full, or after the removal of potential outliers (Pi=0.159; *p*=0.82).

The previous results suggest that visual reward-driven distraction interferes with auditory temporal representations as observers encode the AV scene. This may be a result of weaker stability of responses to attended auditory stimuli (Lakatos et al., 2013) over distraction time, as indexed by single-trial phase locking measures. An alternative interpretation is that in any error trial, not only VDAC, PLVs have not consistently been held throughout the process of AV evidence integration in favor of a correct target. To further investigate this distinction, we examined whether PLV changes directly predicted outcomes across all trials, regardless of condition. This would be consistent with our findings that reward-associated signals have the ability to reduce PLV and that such reductions predict performance changes under reward-driven distraction. We were motivated to address whether spontaneous PLV fluctuations on a trial may directly influence the decision process used to determine choices later at the time of responding. Alternatively, the PLV fluctuations may not impact decisions and outcomes directly. They may instead emerge as a function of other circumstances that do directly impact the decision process and which are more likely under VDAC, including adverse sensory encoding (Hickey & Peelen, 2015) or weak evidence integration (Pearson et al., 2022), for instance.

To distinguish these alternatives, we ran an exploratory analysis designed to determine whether the previously observed statistical interaction between 'Stage' and 'Reward' factors could vary according to the 'Outcome' of a trial (correct or incorrect), i.e., through a triple interaction (see *Methods*). The analysis then consisted of a LME model on the same single-trial PLV data that included the trial factors ‘Stage’, ‘Reward’, and ‘Outcome’ of the result (correct versus incorrect response). If behavioral outcomes relate generally to spontaneous trial-to-trial fluctuations across conditions (PLV as a decision variable scenario), a main effect of ‘Outcome’ is expected. Alternatively, if PLV changes are specific to the decision profile under VDAC, then this may then be supported by a triple interaction. The analysis confirmed the expected interaction between ‘Stage’ and ‘Reward’ ($F_{1,81}$=10.15; $p = 0.002$) as before, and revealed a significant effect of ‘Outcome’ ($F_{1,70}$=6.69; $p = 0.012$) with no significant triple interaction ($F_{1,82}$=0.20; $p = 0.66$). The effect of ‘Outcome’ was due to the fact that, across experimental conditions, trials leading to correct responses revealed greater PLVs than those leading to incorrect outcomes (Figure 3D). No other significant interaction or main effect was observed (Table 6). This indicates that down-modulations on PLV generally predict incorrect trial outcomes, including but not limited to VDAC, suggesting a systematic role for temporally precise auditory neural representations in the AV perceptual decision-making process.

Omnibus F-test with Satterthwaite approximation analysis of single-trial auditory neural locking

Model: PLV ~ Stage*Reward*Outcome + (Stage*Reward*Outcome | participant\difficulty)

| Term | F-value | *p*-value |
|---|---|---|
| Intercept | | |
| Stage | $F_{1,69}$=1.80 | 0.18 |
| Reward | $F_{1,52}$=0.73 | 0.40 |
| Outcome | $F_{1,70}$=6.69 | **0.012** |
| Stage × Reward | $F_{1,81}$=10.15 | **0.002** |
| Stage × Outcome | $F_{1,38}$=0.56 | 0.46 |
| Reward × Outcome | $F_{1,52}$=0.28 | 0.60 |
| Stage × Reward × Outcome | $F_{1,82}$=0.20 | 0.66 |

Values in bold indicate *p*-values below 0.05 significance threshold

**Table 6.** Exploratory re-analysis of auditory PLV peak size data, including Outcome as predictor.

## 4. Discussion

Integrating visual and auditory signals into coherent representations of sensory input matters to enable adaptive interactions with our surroundings. This EEG study found that obsolete reward associations can shape the quality of such representations, limiting the brain’s ability to integrate auditory and visual signals into dynamic multimodal streams. For the first time, we show that value-associated signals hinder the integration of auditory and visual elements of the scene through common temporal cues.

Our results demonstrate that after observers trained on color-reward associations, they became less successful in matching a pulsating sound to one of two visual flickering items when surrounded peripherally by irrelevant but reward-associated color features. Decreased stability of evoked auditory responses was correlated with performance changes. A key neural variable, the quality of auditory locking or temporal precision, predicted trial outcomes across all experiment conditions and was affected by VDAC. We suggest that, in busy scenes, the signaling of reward may compete with temporal processes of dynamic multimodal object integration. The interpretation is that reward-associated distractors may, in this case, impose additional attentional demands that disrupt binding guided by temporal coherence.

*4.1 Impact of value-driven gaps of attention on the dynamics of multimodal integration.* Behavioral reports of cross-modal VDAC hint at the possibility that learned value associations reorganize multimodal processing. In one study (Sanz et al., 2018), irrelevant but value-associated sounds delayed performance on a visual target detection task when sound and image elements were semantically related. By comparison, speed was unaffected when semantic relationships were lacking. In another inter-modal competition study, the effects of sound distractors on arbitrary (not semantically related) visual targets were addressed (Anderson, 2016). In this AV adaptation of the classical VDAC visual paradigm (Anderson et al., 2011), participants’ attention was generally diverted away from visual targets when reward-associated sound distractors were present.  also showed a cross-modal VDAC design but found opposite results, as in their study, the presence of reward-associated auditory stimuli improved visuospatial performance instead. The differing results across these studies may relate to variations in the amount of competition that is needed to perform the task correctly. Indeed, mixed distraction/facilitation behavioral results can be observed across VDAC studies (Rusz et al., 2020). In the case of Pooresmaeili and colleagues, AV displays were systematically controlled for spatial congruence, and facilitation findings were also observed in similar congruence designs (Grahek et al., 2021). When reward cues do not spatially match the target, capture may turn into an attentional cost or distraction (Tankelevitch et al., 2020).

Neuroimaging VDAC studies evidence response changes in the sensory cortex corresponding to the value-associated distractor's modality (Itthipuripat et al., 2019; A. J. Kim et al., 2021). Some of these findings suggest that sensory areas produce more finely tuned representations of the reward-associated distractors, perhaps analogously to the effects of top-down attentional selection, and at similar time scales (Stănişor et al., 2013). Distractors that happen to signal relatively less value by contrast will tend to have a relatively lower degree of representation and be easier to ignore. The timeline for phasic modulations by VDAC differs according to whether the distractor is from the same target modality (beginning around 100 ms, for example (MacLean & Giesbrecht, 2015), but see (Tankelevitch et al., 2020)) or elsewhere (established around 200 ms) (Vakhrushev et al., 2023). Nevertheless, these reports agree on a relatively early window for VDAC influences to be observed cross-modally rather than through top-down processing (Cervantes Constantino et al., 2023). Loss of precision effects is expected to be stronger for multisensory scenes where reward-associated signals are inter-modal (i.e., reach sensory networks related to current task goals from a different sensory area) as opposed to intra-modal. How the sensory path traveled by VDAC signals constrains multimodal latency and phase relationships in continuous dynamic conditions offers an interesting avenue of research for future studies.

*4.2 VDAC alters stimulus-driven brain dynamics that are required for behavior.* To link information within and across diverse area networks in the brain, their dynamics of excitability may be adjusted so they can be shared (Bauer et al., 2020; Fries, 2015; Keil & Senkowski, 2018). In unimodal and multimodal networks, this creates oscillatory phase relationships that can be exploited to distinguish distractors from targets (Lakatos et al., 2013; Nobre & van Ede, 2023). The adjustment mechanism may be exerted by top-down attention, which orchestrates the phase reset and entrainment of neural fluctuations that are tuned to the target stream (Lakatos et al., 2008). We propose a similar adjustment mechanism by VDAC, where excitability modulations lead instead to relative representational gains for the reward-associated distractor at the expense of coherent auditory locking. This predicts that distractors are processed as independent elements from the auditory stream (Figure 4), avoiding or preventing multimodal integration processes. Underscoring this proposed mechanism is our finding that reward-associated distractors failed to support cross-modal association despite being dynamically co-modulated with sound themselves. Under voluntary top-down attention, existing temporal correspondences would typically boost the precision of auditory responses to temporal patterns (Kayser, 2009). Yet through VDAC, the reliability of stimulus-driven activity in the AC appeared to be reduced. These changes also corresponded with a lack of improvement in behavioral performance over trials, consistent with the weakening of neural representations of targets by salient distractors (Hickey & Peelen, 2015). Our findings hence support the principle that engaging with reward-associated cues in the environment impacts robust encoding of relevant sensory-driven dynamics. This provides insight into the temporal dimension of VDAC in more naturalistic conditions.

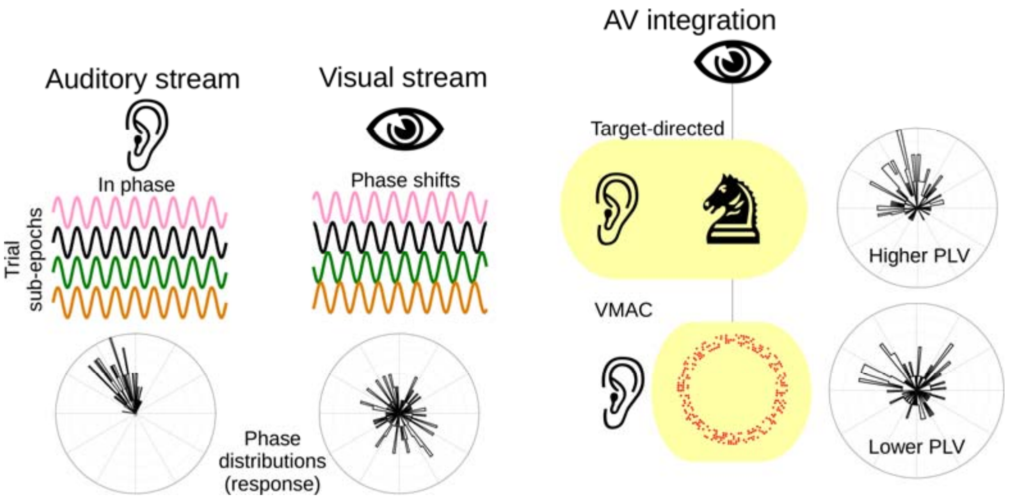


**Figure 4**. **Illustration of a conceptual model to explain the cross-modal dynamics of reward-driven distraction in the task.** During stimulation, unimodal responses to the auditory stream can be sub-epoched

(left), and these segments will be maximally coherent with each other as their periodicity is determined by synchronizing to the AM sound pattern's fixed rate. Phase diagrams below represent the physiological responses with noise, resulting in a relatively compact phase-locking distribution for unimodal auditory responses. For unimodal visual responses (middle), the dual visual stream (2-flicker presentation) imposes systematic shifts of attentional sampling behavior among the two candidate choices, resulting in decreased synchronization to the particular target rate. The unimodal visual contribution to responses locked to this target rate is hence represented by relatively lower phase coherence. In a successful, target-directed AV integration trial (right), participants optimize their behavior by binding auditory and target visual streams over time, i.e., maximizing the stability of the constant inter-modal phase correspondence. Auditory responses to the AV scenes may maintain more precise PLVs proportionally to the periods of sampling stability. By contrast, for a VDAC unsuccessful trial, sampling of peripheral reward-associated distractors attracts attention, thereby reducing precision. Critically, the distractor is here synced to the visual target and also flickers at the auditory rate, indicating that unimodal auditory representations are segregated from dynamic value signals under VDAC.

*4.3 Reward-driven salience modulations on endogenous excitability rhythms and cross-modal interactions.* Our paradigm concerns sensory-driven rather than endogenous oscillation dynamics. Endogenous alpha/beta band cortical synchronization can be referenced to internal timing mechanisms driven by thalamocortical pacemaker circuits (Fries, 2015; Keitel et al., 2019; Lakatos et al., 2016) to mediate top-down attentional influence. Stimulus-driven band activity, on the other hand, is subject to control by top-down attention (Gray et al., 2015; Haegens & Zion Golumbic, 2018; Y. J. Kim et al., 2007) and, as shown in our study, modulated by VDAC. Habitual biases of VDAC appear more driven by the updating involved in the computation of relative value and its consolidation (Anderson, 2019; S. Kim & Beck, 2020; Liao & Anderson, 2020) rather than amplifying sensory features that signal value (Anderson et al., 2021). In addition, the redress by VDAC may involve object-selective cortical processing areas (Anderson et al., 2021; Hickey & Peelen, 2015), with the encoding of relative value in object-based representations (Stănişor et al., 2013). In this way, endogenous band activity corresponding to value-associated distractors may potentially encode input signals in predominantly different representational formats than that of its feedforward temporal dynamics. Notably, gamma-band activity varies across different regions of a sensory area, with representations of bottom-up stimulus salience, stimulus history, and top-down influences (Fries, 2015). By this account, the endogenous alpha/beta rhythms would phase re-set local gamma rhythms corresponding to the salient peripheral locations, to the effect of tuning and strengthening them on account of their value representations. Local visual endogenous alpha rhythms modulate the cortical representation of stimuli, with band activity increments associated with suppression of dedicated processing to irrelevant background regions (Gray et al., 2015). Beta rhythms may be synergistically involved by releasing from such dedicated suppression, allowing tracking of input dynamics at attended locations (Lee et al., 2013).

Neuroimaging suggests that VDAC outcompetes top-down visuospatial attention through the release of inhibition of visual cortical regions that encode salient distractors (Anderson et al., 2021; Wang et al., 2015). Top-down visuospatial and feature-based influences on visual cortical processing (including early visual cortex, extrastriate, and ventral visual cortex) are mediated by fronto-parietal attentional networks; by comparison, VDAC is mediated by the plasticity of stimulus representations across networks that include posterior parietal cortical (PPC) areas and the frontal eye field (Anderson, 2019). Furthermore, nuclei such as the tail of the caudate nucleus, substantia nigra, superior colliculus, and amygdala are involved in the rapid orienting of VDAC and modulate visual cortical processing through thalamic connections (Anderson et al., 2021; Tankelevitch et al., 2020).

A candidate origin for these distributed VDAC biases may be the locus coeruleus (LC)-related functional networks (Todd & Manaligod, 2018), since this system is quickly sensitive to changes in stimulus significance and mediates attentional shifts to optimize behavior (Sara, 2009). The LC system also modulates high-frequency oscillations (such as gamma) to favor the transmission of information related to prioritized distractor representations across neuronal ensembles (Mather et al., 2016). Salience-guided LC influences on visual encoding may then be mediated by networks that include the frontal lobe and amygdala. Notably, the amygdala projects input to spatial priority maps in the posterior parietal cortex (PPC) that predict where saccades will be directed (Bisley & Goldberg, 2010; Markovic et al., 2014). These priority maps appear to be a crucial feature of salience encoding in VDAC, as PPC responses respond selectively to learned high-value distractors (A. J. Kim et al., 2021; Peck et al., 2009; Tankelevitch et al., 2020). Moreover, the PPC, and intraparietal sulcus in particular, participate in pathways for cross-modal communication and auditory-to-visual entrainment. These pathways extend to the prefrontal and superior temporal cortices, as well as subcortical relays (superior colliculus and thalamus)  (Bauer et al., 2020, 2021; Lakatos et al., 2009). In a cross-modal VDAC study, superior temporal cortex responses were sensitive to reward-association contingencies from sound distractors and predictive of behavioral outcomes, as they were in target-related visual cortex (Pooresmaeili et al., 2014).

### *4.4 Reward-driven distraction under rhythmic attentional sampling: a dynamical framework.*

Priority maps appear computationally relevant for dynamic scenarios because of their flexibility to reconfigure in accordance with what has been previously relevant in similar situations. Dynamic fluctuations by VDAC may follow in the frame of browsing the overall space of such context-dependent possibilities in priority maps, for instance, on a theta-based rhythmic sampling process (Lakatos et al., 2016; Landau & Fries, 2012; Todd & Manaligod, 2018). Under this principle, preferential sampling does not mean continuous sampling at all times. Rather, regular periods of disengagement, referred to as ‘down’ phases of the attentional rhythmic cycle, allow exploratory shifts in attention. These periodic attenuations create windows of opportunity to reweight the priority map (Fiebelkorn & Kastner, 2019). VDAC may seize on these temporary blind spots, boosting the chance of exploratory oculomotor behavior (Rusz et al., 2020).

As a working model, in the following, we provide one mechanistic framework of the rhythmic sampling account applicable to VDAC in the present study. First, a distractor successfully draws attention away from the target space, defining the time when regions processing distractor features receive an ‘attentional spotlight’. The term refers to the establishment of a preferred phase in the oscillatory excitability patterns that involve regions entrained by the attended distractor. Crucially, the spotlight is part of a multidimensional phase excitability map parameter, established spatiotemporally (and on higher-order multidimensional feature space) (Lakatos et al., 2019). *Second*, setting this new VDAC phase will be determined by not only the pace of the distractor’s flicker but also its spatial and featural domain organization. For example, the theta sampling selection strategy may organize and discriminate phases (Fiebelkorn & Kastner, 2019; Landau & Fries, 2012) assigned to left versus right hemifields, centre versus surround regions, red versus grey hues, and so on. This prevents the temporal flicker phase from univocally determining the spotlight. *Third*, elements not attended at the time of capture, namely the VDAC 'background', become attenuated. One reason concerns those elements that were following alternative dynamics from the value-associated distractor's (e.g., grey dots and the incorrect chess piece option in our study). Attenuation happens there simply because they are out of sync with the attended domain. Any stimulus encoding relating to these parts of the background will generally be fed forward with random phase excitability levels, dictating a poorer signal-to-noise ratio beyond local networks. *Fourth*, parts of the VDAC background that do follow the same dynamics as the attended distractors (in our study, the correct visual target and the sound's AM) are not enhanced. The reason is that their feed-forward processing is relatively less stable, stemming from their failure to consistently match the high excitability phase of the distractor's multi-dimensional feature spotlight. That optimal phase changes as the current phase excitability map warps on attending in and out of the distractor content (Lakatos et al., 2019). In this way, reward-driven distraction may impact the temporal organization of a busy scene through the exploitation of gaps in the top-down rhythmic sampling strategy, followed by later adjustments. When busy scenes yield potential for strong multimodal object formation, reward-associated signals may disrupt otherwise sustainable representations of key temporal correspondences for multisensory binding.

## 5. Data and code availability.

Data and code supporting the findings of this study are openly available at the Open Science Framework at http://doi.org/10.17605/OSF.IO/SPJ9F.

## 6. Author contributions

**Rodrigo Caramés Harcevnicow:** Formal analysis, Writing - Original Draft, Writing - Review & Editing, Visualization **Thaiz Sánchez-Costa:** Investigation, Data Curation, Writing - Review & Editing **Alejandra Carboni:** Software, Resources, Writing - Review & Editing, Supervision, Funding acquisition **Francisco Cervantes Constantino:** Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data Curation, Writing - Original Draft, Writing - Review & Editing, Visualization, Supervision, Project administration, Funding acquisition

## 7. Funding

The research presented in this manuscript received funding from the Agencia Nacional de Investigación e Innovación, Uruguay, under the code PD_NAC_2018_1_150365.

## 8. Declaration of Competing Interests

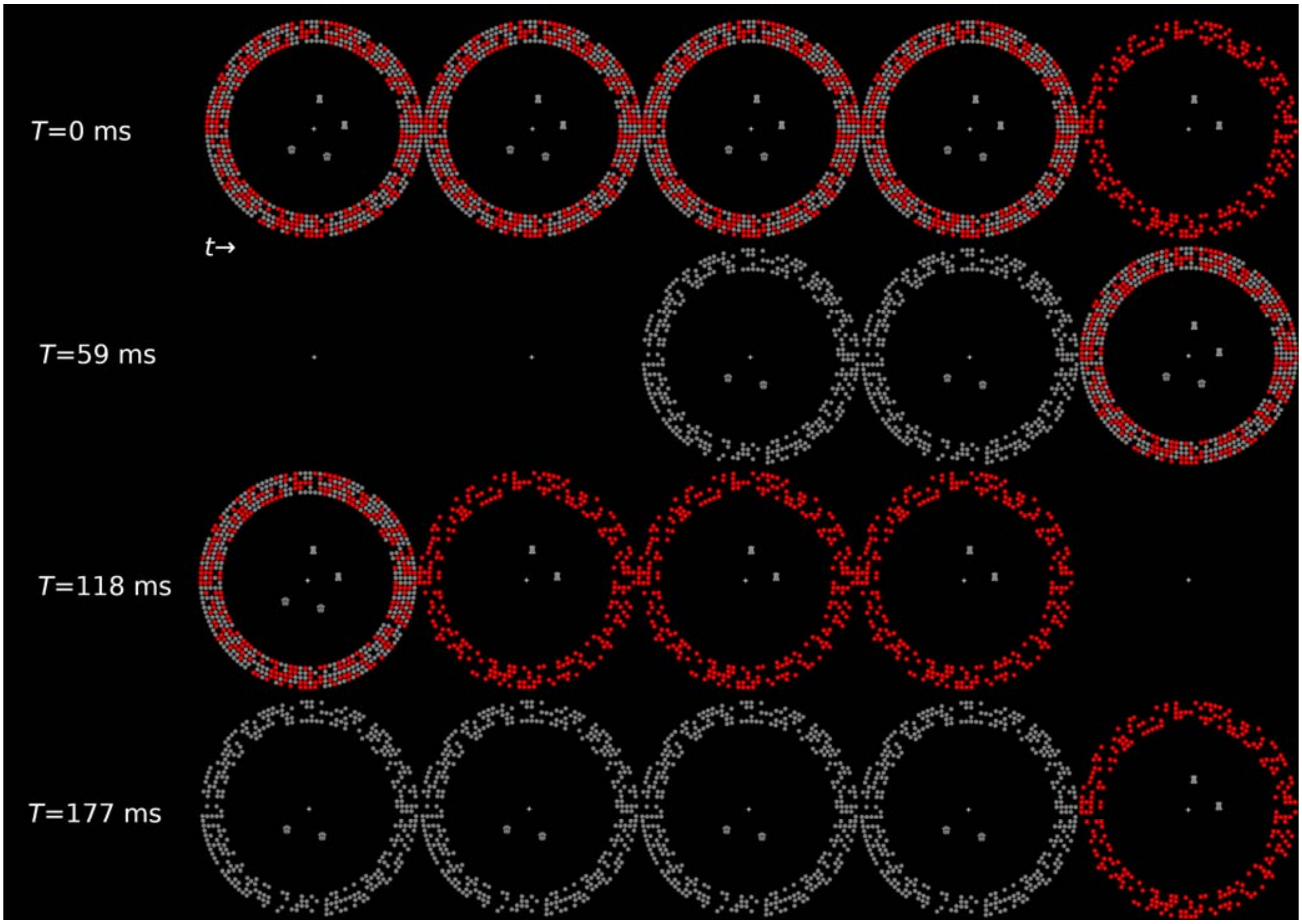
None.

## 9. Acknowledgments

We thank Alex L. White and Germán A. Cipriani for code sharing, and Franco Caballero and Cecilia Biurrun for assistance with data collection.

## 10. Supplementary Material

**Video 1**. Sample AV testing (Tt1/Tt2) trial with a 9 Hz AM sound presentation. Both central figures shift locations every 1 second in this example. See Supplementary Fig. 1 for a video frame detail of the first two cycles of the target rate.



**Supplementary Figure 1**. Frame detail of the first 220 ms of Video 1. Time *t* step increases correspond to the 12 ms frame period approximately. Due to the frequency approximation approach (see *Methods*), the duty cycle of the 9 Hz set is 19 frames, and 15 frames for the 11 Hz set. The correct figure in every trial (rook in this example) syncs exclusively with the colored dot distractor set.

## 11. References

Anderson, B. A. (2016). Value-driven attentional capture in the auditory domain. *Attention, Perception & Psychophysics*, *78*(1), 242–250. https://doi.org/10.3758/s13414-015-1001-7

Anderson, B. A. (2019). Neurobiology of value-driven attention. *Current Opinion in Psychology*, *29*, 27–33. https://doi.org/10.1016/j.copsyc.2018.11.004

Anderson, B. A., Kim, H., Kim, A. J., Liao, M.-R., Mrkonja, L., Clement, A., & Grégoire, L. (2021). The past, present, and future of selection history. *Neuroscience & Biobehavioral Reviews*, *130*, 326–350. https://doi.org/10.1016/j.neubiorev.2021.09.004

Anderson, B. A., Laurent, P. A., & Yantis, S. (2011). Value-driven attentional capture. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(25), 10367–10371. https://doi.org/10.1073/pnas.1104047108

Atilgan, H., Town, S. M., Wood, K. C., Jones, G. P., Maddox, R. K., Lee, A. K. C., & Bizley, J. K. (2018). Integration of Visual Information in Auditory Cortex Promotes Auditory Scene Analysis through Multisensory Binding. *Neuron*, *97*(3), 640-655.e4. https://doi.org/10.1016/j.neuron.2017.12.034

Bauer, A.-K. R., Debener, S., & Nobre, A. C. (2020). Synchronisation of Neural Oscillations and Cross-modal Influences. *Trends in Cognitive Sciences*, *24*(6), 481–495. https://doi.org/10.1016/j.tics.2020.03.003

Bauer, A.-K. R., Ede, F. van, Quinn, A. J., & Nobre, A. C. (2021). Rhythmic Modulation of Visual Perception by Continuous Rhythmic Auditory Stimulation. *Journal of Neuroscience*, *41*(33), 7065–7075. https://doi.org/10.1523/JNEUROSCI.2980-20.2021

Bisley, J. W., & Goldberg, M. E. (2010). Attention, intention, and priority in the parietal lobe. *Annual Review of Neuroscience*, *33*, 1–21. https://doi.org/10.1146/annurev-neuro-060909-152823

Bizley, J. K., Maddox, R. K., & Lee, A. K. C. (2016). Defining Auditory-Visual Objects: Behavioral Tests and Physiological Mechanisms. *Trends in Neurosciences*, *39*(2), 74–85. https://doi.org/10.1016/j.tins.2015.12.007

Cervantes Constantino, F., Sánchez-Costa, T., Cipriani, G. A., & Carboni, A. (2023). Visuospatial attention revamps cortical processing of sound amid audiovisual uncertainty. *Psychophysiology*, e14329. https://doi.org/10.1111/psyp.14329

Chen, G., & Gong, P. (2022). A spatiotemporal mechanism of visual attention: Superdiffusive motion and theta oscillations of neural population activity patterns. *Science Advances*, *8*(16), eabl4995. https://doi.org/10.1126/sciadv.abl4995

Cho, S. A., & Cho, Y. S. (2021). Uncertainty modulates value-driven attentional capture. *Attention, Perception, & Psychophysics*, *83*(1), 142–155. https://doi.org/10.3758/s13414-020-02171-3

de Cheveigné, A., & Simon, J. Z. (2007). Denoising based on time-shift PCA. *Journal of Neuroscience Methods*, *165*(2), 297–305. https://doi.org/10.1016/j.jneumeth.2007.06.003

de Cheveigné, A., & Simon, J. Z. (2008). Sensor noise suppression. *Journal of Neuroscience Methods*, *168*(1), 195–202. https://doi.org/10.1016/j.jneumeth.2007.09.012

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics*, *16*(1), 143–149. https://doi.org/10.3758/BF03203267

Fiebelkorn, I. C., & Kastner, S. (2019). A Rhythmic Theory of Attention. *Trends in Cognitive Sciences*, *23*(2), 87–101. https://doi.org/10.1016/j.tics.2018.11.009

Fries, P. (2015). Rhythms For Cognition: Communication Through Coherence. *Neuron*, *88*(1), 220–235. https://doi.org/10.1016/j.neuron.2015.09.034

Grahek, I., Schettino, A., Koster, E. H. W., & Andersen, S. K. (2021). Dynamic Interplay between Reward and Voluntary Attention Determines Stimulus Processing in Visual Cortex. *Journal of Cognitive Neuroscience*, *33*(11), 2357–2371. https://doi.org/10.1162/jocn_a_01762

Gray, M. J., Frey, H.-P., Wilson, T. J., & Foxe, J. J. (2015). Oscillatory recruitment of bilateral visual cortex during spatial attention to competing rhythmic inputs. *Journal of Neuroscience*, *35*(14), 5489–5503. Scopus. https://doi.org/10.1523/JNEUROSCI.2891-14.2015

Haegens, S., & Zion Golumbic, E. (2018). Rhythmic facilitation of sensory processing: A critical review. *Neuroscience and Biobehavioral Reviews*, *86*, 150–165. https://doi.org/10.1016/j.neubiorev.2017.12.002

Hickey, C., & Peelen, M. V. (2015). Neural mechanisms of incentive salience in naturalistic human vision. *Neuron*, *85*(3), 512–518. https://doi.org/10.1016/j.neuron.2014.12.049

Hyvarinen, A. (1999). Fast and robust fixed-point algorithms for independent component analysis. *IEEE Transactions on Neural Networks*, *10*(3), 626–634. https://doi.org/10.1109/72.761722

Itthipuripat, S., Vo, V. A., Sprague, T. C., & Serences, J. T. (2019). Value-driven attentional capture enhances distractor representations in early visual cortex. *PLOS Biology*, *17*(8), e3000186. https://doi.org/10.1371/journal.pbio.3000186

Jenkins, J., Rhone, A. E., Idsardi, W. J., Simon, J. Z., & Poeppel, D. (2011). The Elicitation of Audiovisual Steady-State Responses: Multi-Sensory Signal Congruity and Phase Effects. *Brain Topography*, *24*(2), 134–148. https://doi.org/10.1007/s10548-011-0174-1

Junghöfer, M., Elbert, T., Tucker, D. M., & Rockstroh, B. (2000). Statistical control of artifacts in dense array EEG/MEG studies. *Psychophysiology*, *37*(04), 523–532. https://doi.org/null

Kayser, C. (2009). Phase Resetting as a Mechanism for Supramodal Attentional Control. *Neuron*, *64*(3), 300–302. https://doi.org/10.1016/j.neuron.2009.10.022

Keil, J., & Senkowski, D. (2018). Neural Oscillations Orchestrate Multisensory Processing. *The Neuroscientist*, *24*(6), 609–626. https://doi.org/10.1177/1073858418755352

Keitel, C., Keitel, A., Benwell, C. S. Y., Daube, C., Thut, G., & Gross, J. (2019). Stimulus-Driven Brain Rhythms within the Alpha Band: The Attentional-Modulation Conundrum. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *39*(16), 3119–3129. https://doi.org/10.1523/JNEUROSCI.1633-18.2019

Keitel, C., & Müller, M. M. (2016). Audio-visual synchrony and feature-selective attention co-amplify early visual processing. *Experimental Brain Research*, *234*(5), 1221–1231. https://doi.org/10.1007/s00221-015-4392-8

Kim, A. J., Grégoire, L., & Anderson, B. A. (2021). Value-Biased Competition in the Auditory System of the Brain. *Journal of Cognitive Neuroscience*, *34*(1), 180–191. https://doi.org/10.1162/jocn_a_01785

Kim, S., & Beck, M. R. (2020). Impact of relative and absolute values on selective attention. *Psychonomic Bulletin & Review*, *27*(4), 735–741. https://doi.org/10.3758/s13423-020-01729-4

Kim, Y. J., Grabowecky, M., Paller, K. A., Muthu, K., & Suzuki, S. (2007). Attention induces synchronization-based response gain in steady-state visual evoked potentials. *Nature Neuroscience*, *10*(1), 117–125. https://doi.org/10.1038/nn1821

Koenig, S., Kadel, H., Uengoer, M., Schubö, A., & Lachnit, H. (2017). Reward Draws the Eye, Uncertainty Holds the Eye: Associative Learning Modulates Distractor Interference in Visual Search. *Frontiers in Behavioral Neuroscience*, *11*. https://www.frontiersin.org/articles/10.3389/fnbeh.2017.00128

Kösem, A., Gramfort, A., & van Wassenhove, V. (2014). Encoding of event timing in the phase of neural oscillations. *NeuroImage*, *92*, 274–284. https://doi.org/10.1016/j.neuroimage.2014.02.010

Lakatos, P., Barczak, A., Neymotin, S. A., McGinnis, T., Ross, D., Javitt, D. C., & O'Connell, M. N. (2016). Global dynamics of selective attention and its lapses in primary auditory cortex. *Nature Neuroscience*, *19*(12), 1707–1717. https://doi.org/10.1038/nn.4386

Lakatos, P., Gross, J., & Thut, G. (2019). A New Unifying Account of the Roles of Neuronal Entrainment. *Current Biology*, *29*(18), R890–R905. https://doi.org/10.1016/j.cub.2019.07.075

Lakatos, P., Karmos, G., Mehta, A. D., Ulbert, I., & Schroeder, C. E. (2008). Entrainment of neuronal oscillations as a mechanism of attentional selection. *Science (New York, N.Y.)*, *320*(5872), 110–113. https://doi.org/10.1126/science.1154735

Lakatos, P., Musacchia, G., O'Connell, M. N., Falchier, A. Y., Javitt, D. C., & Schroeder, C. E. (2013). The spectrotemporal filter mechanism of auditory selective attention. *Neuron*, *77*(4), 750–761. https://doi.org/10.1016/j.neuron.2012.11.034

Lakatos, P., O'Connell, M. N., Barczak, A., Mills, A., Javitt, D. C., & Schroeder, C. E. (2009). The leading sense: Supramodal control of neurophysiological context by attention. *Neuron*, *64*(3), 419–430. https://doi.org/10.1016/j.neuron.2009.10.014

Landau, A. N., & Fries, P. (2012). Attention samples stimuli rhythmically. *Current Biology: CB*, *22*(11), 1000–1004. https://doi.org/10.1016/j.cub.2012.03.054

Lee, J. H., Whittington, M. A., & Kopell, N. J. (2013). Top-Down Beta Rhythms Support Selective Attention via Interlaminar Interaction: A Model. *PLOS Computational Biology*, *9*(8), e1003164. https://doi.org/10.1371/journal.pcbi.1003164

Liao, M.-R., & Anderson, B. A. (2020). Inertia in value-driven attention. *Learning & Memory*, *27*(12), 488–492. https://doi.org/10.1101/lm.052027.120

MacLean, M. H., & Giesbrecht, B. (2015). Neural evidence reveals the rapid effects of reward history on selective attention. *Brain Research*, *1606*, 86–94. https://doi.org/10.1016/j.brainres.2015.02.016

Maddox, R. K., Atilgan, H., Bizley, J. K., & Lee, A. K. C. (2015). Auditory selective attention is enhanced by a task-irrelevant temporally coherent visual stimulus in human listeners. *eLife*, *4*. https://doi.org/10.7554/eLife.04995

Markovic, J., Anderson, A. K., & Todd, R. M. (2014). Tuning to the significant: Neural and genetic processes underlying affective enhancement of visual perception and memory. *Behavioural Brain Research*, *259*, 229–241. https://doi.org/10.1016/j.bbr.2013.11.018

Mather, M., Clewett, D., Sakaki, M., & Harley, C. W. (2016). Norepinephrine ignites local hotspots of neuronal excitation: How arousal amplifies selectivity in perception and memory. *Behavioral and Brain Sciences*, *39*, e200. https://doi.org/10.1017/S0140525X15000667

Monosov, I. E. (2020). How Outcome Uncertainty Mediates Attention, Learning, and Decision-Making. *Trends in Neurosciences*, *43*(10), 795–809. https://doi.org/10.1016/j.tins.2020.06.009

Nakanishi, M., Wang, Y., Wang, Y.-T., Mitsukura, Y., & Jung, T.-P. (2014). Generating Visual Flickers for Eliciting Robust Steady-State Visual Evoked Potentials at Flexible Frequencies Using Monitor Refresh Rate. *PLOS ONE*, *9*(6), e99235. https://doi.org/10.1371/journal.pone.0099235

Nobre, A. C., & van Ede, F. (2023). Attention in flux. *Neuron*, *111*(7), 971–986. https://doi.org/10.1016/j.neuron.2023.02.032

Nozaradan, S., Peretz, I., & Mouraux, A. (2012). Steady-state evoked potentials as an index of multisensory temporal binding. *NeuroImage*, *60*(1), 21–28. https://doi.org/10.1016/j.neuroimage.2011.11.065

Pearson, D., Watson, P., Albertella, L., & Le Pelley, M. E. (2022). Attentional economics links value-modulated attentional capture and decision-making. *Nature Reviews Psychology*, *1*(6), Article 6. https://doi.org/10.1038/s44159-022-00053-z

Peck, C. J., Jangraw, D. C., Suzuki, M., Efem, R., & Gottlieb, J. (2009). Reward modulates attention independently of action value in posterior parietal cortex. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *29*(36), 11182–11191. https://doi.org/10.1523/JNEUROSCI.1929-09.2009

Peirce, J. W. (2007). PsychoPy—Psychophysics software in Python. *Journal of Neuroscience Methods*, *162*(1), 8–13. https://doi.org/10.1016/j.jneumeth.2006.11.017

Pernet, C., Wilcox, R., & Rousselet, G. (2013). Robust Correlation Analyses: False Positive and Power Validation Using a New Open Source Matlab Toolbox. *Frontiers in Psychology*, *3*. https://www.frontiersin.org/articles/10.3389/fpsyg.2012.00606

Pooresmaeili, A., FitzGerald, T. H. B., Bach, D. R., Toelch, U., Ostendorf, F., & Dolan, R. J. (2014). Cross-modal effects of value on perceptual acuity and stimulus encoding. *Proceedings of the National Academy of Sciences*, *111*(42), 15244–15249. https://doi.org/10.1073/pnas.1408873111

Rolls, E. T., Deco, G., Huang, C.-C., & Feng, J. (2022). The human language effective connectome. *NeuroImage*, *258*, 119352. https://doi.org/10.1016/j.neuroimage.2022.119352

Rouhani, N., Norman, K. A., & Niv, Y. (2018). Dissociable effects of surprising rewards on learning and memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*(9), 1430–1443. https://doi.org/10.1037/xlm0000518

Rusz, D., Le Pelley, M. E., Kompier, M. A. J., Mait, L., & Bijleveld, E. (2020). Reward-driven distraction: A meta-analysis. *Psychological Bulletin*, *146*(10), 872–899. https://doi.org/10.1037/bul0000296

Sanz, L. R. D., Vuilleumier, P., & Bourgeois, A. (2018). Cross-modal integration during value-driven attentional capture. *Neuropsychologia*, *120*, 105–112. https://doi.org/10.1016/j.neuropsychologia.2018.10.014

Sara, S. J. (2009). The locus coeruleus and noradrenergic modulation of cognition. *Nature Reviews Neuroscience*, *10*(3), Article 3. https://doi.org/10.1038/nrn2573

Schwarzkopf, D., de Haas, B., & Rees, G. (2012). Better Ways to Improve Standards in Brain-Behavior Correlation Analysis. *Frontiers in Human Neuroscience, 6*. https://www.frontiersin.org/articles/10.3389/fnhum.2012.00200

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers*, *31*(1), 137–149. https://doi.org/10.3758/BF03207704

Stănişor, L., van der Togt, C., Pennartz, C. M. A., & Roelfsema, P. R. (2013). A unified selection signal for attention and reward in primary visual cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(22), 9136–9141. https://doi.org/10.1073/pnas.1300117110

Tankelevitch, L., Spaak, E., Rushworth, M. F. S., & Stokes, M. G. (2020). Previously Reward-Associated Stimuli Capture Spatial Attention in the Absence of Changes in the Corresponding Sensory Representations as Measured with MEG. *Journal of Neuroscience*, *40*(26), 5033–5050. https://doi.org/10.1523/JNEUROSCI.1172-19.2020

Todd, R. M., & Manaligod, M. G. M. (2018). Implicit guidance of attention: The priority state space framework. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *102*, 121–138. https://doi.org/10.1016/j.cortex.2017.08.001

Vakhrushev, R., Cheng, F. P.-H., Schacht, A., & Pooresmaeili, A. (2023). Differential effects of intra-modal and cross-modal reward value on perception: ERP evidence. *PLOS ONE*, *18*(6), e0287900. https://doi.org/10.1371/journal.pone.0287900

van Diepen, R. M., & Mazaheri, A. (2018). The Caveats of observing Inter-Trial Phase-Coherence in Cognitive Neuroscience. *Scientific Reports*, *8*(1), 2990. https://doi.org/10.1038/s41598-018-20423-z

Wang, L., Yu, H., Hu, J., Theeuwes, J., Gong, X., Xiang, Y., Jiang, C., & Zhou, X. (2015). Reward breaks through center-surround inhibition via anterior insula. *Human Brain Mapping*, *36*(12), 5233–5251. https://doi.org/10.1002/hbm.23004