Detecting behavioural oscillations with increased sensitivity:

A modification of Brookshire's (2022) AR-surrogate method

Anthony M. Harris[1#] & Henry A. Beale[1]

[1] Queensland Brain Institute, The University of Queensland, St Lucia, Australia, 4072

[#] Corresponding author

This work was funded by an ARC Discovery Early Career Researcher Award awarded to AMH (DE220101019).

Correspondence: anth.m.harris@gmail.com

The authors declare no competing interests.

Abstract

A core challenge of cognitive neuroscience is to understand how cognition changes over time within the same individual. For example, the tendency for behavioural responses in a range of cognitive domains to oscillate over time has been studied extensively. Recently, however, the phenomenon of behavioural oscillations has been called into question by indications that past findings might reflect aperiodic temporal structure rather than true oscillations. Brookshire (2022) proposed methods to control for aperiodic temporal structure while examining oscillations in behavioural time-courses and found no evidence of behavioural oscillations in reanalyses of four published datasets. However, Brookshire's (2022) method has been criticised for having low sensitivity to detect effects of realistic magnitude, so it is currently unclear whether these findings suggest that behavioural oscillations are not present in these and perhaps many other datasets, or whether they are false negatives. Here, we present a modification of Brookshire's (2022) AR-surrogate method with increased sensitivity to detect effects of realistic magnitude, adequate control of false positives, and other desirable properties such as the ability to increase statistical power by adding more participants. Using this method, we reanalyse the same publicly available datasets and show significant behavioural oscillations in each of them, suggesting oscillations in behaviour are a robust phenomenon upon which to draw theoretical inferences. The participant-level AR-surrogate method is currently the most sensitive method available for analysing behavioural oscillations while controlling for the contribution of aperiodic data fluctuations.

In contrast to our seemingly continuous experience of the world, a wealth of research has shown that the quality of perception and other cognitive functions fluctuates rhythmically across brief time scales (Keitel et al., 2022; Kienitz et al., 2022). For example, it has been shown that the probability of detecting a brief visual contrast transient fluctuates between better and worse detection at a rate of around 4 Hz in the second following a high-contrast 'reset event' (Fiebelkorn et al., 2013; Landau & Fries, 2012). Rhythmicity has been shown in a range of psychological phenomena including contrast sensitivity, phosphene perception, attentional modulation, eye-movement timing, working memory performance and many others (Dugué et al., 2016; Hogendoorn, 2016; Pomper & Ansorge, 2021; Romei et al., 2012). Such findings have been used to support influential rhythmic theories of neural processing, in which neural communication and processing are organised by rhythmic brain activity (Fries, 2015, 2023). Recently, however, criticisms have been levelled at the typical analysis methods employed in this literature. Brookshire (2022) argued convincingly that the typical 'shuffling in time' methods employed in this literature test the null hypothesis that there is no temporal structure to the data, rather than testing for the presence of rhythmicity per se. Thus, results produced by these methods may reflect aperiodic temporal structure due to alerting, inhibition of return, and so on, rather than the presence of true 'behavioural oscillations.' Brookshire (2022) proposed two alternative tests that directly target rhythmicity in behavioural time-courses while controlling for the influence of aperiodic trends in the data. Using these methods, Brookshire reanalysed several publicly available datasets and found no evidence for behavioural oscillations, thus casting doubt on the existence of behavioural oscillations as a phenomenon. Here we present a modification of Brookshire's AR surrogate analysis that provides increased sensitivity to detect behavioural oscillations at the group level. We show this method recovers significant behavioural oscillations in the same publicly available datasets analysed by Brookshire (2022), confirming the existence of rhythmicity in behavioural time-courses.

Interest in rhythmic properties of perception has a long and rich history in psychology and philosophy (for excellent reviews, see Kienitz et al., 2022; Menétrey et al., 2022). Modern interest in behavioural oscillations grew out of findings that the phase of rhythmic brain activity (i.e., whether brain waves are at a peak or trough) influenced neural processing and perception (Busch et al., 2009; Dugué et al., 2011; Haegens et al., 2011; Harris et al., 2018; Mathewson et al., 2009; Williams et al., 2024). Following these findings, researchers reasoned that if neural oscillations could be aligned across different trials, such as through a phase reset, then similar phasic fluctuations should be observed in the time-course of behaviour. Early authors tested this idea by presenting a high-contrast 'reset stimulus' followed after a variable interval by a threshold-level target. They found that the detection rate for such targets oscillated rhythmically in the wake of a reset event (Fiebelkorn et al., 2013; Landau & Fries, 2012; Song et al., 2014), consistent with the role of neural oscillations in stimulus processing observed with EEG.

The typical approach to assessing rhythmicity in behavioural time-courses has been to compare the group-average behavioural time-course to a null distribution created by shuffling the mapping of behavioural responses to time-points to create a series of surrogate time-courses in which no temporal structure is present (Fiebelkorn et al., 2013; Landau & Fries, 2012; Tosato et al., 2022). The amplitude spectrum produced by Fourier transforming the true behavioural data is then compared to the distribution of null spectra produced by the surrogate time-courses. The resulting p-value is the proportion of the null spectra that exceeded the amplitude of the true spectrum. This p-value is then corrected for multiple comparisons across frequencies and, if $p < .05$, is interpreted as evidence that oscillations are present in the pattern of behavioural data.

A concern with any application of Fourier approaches is that aperiodic (non-oscillatory) trends in the data will influence the shape of the overall amplitude spectrum, with oscillations appearing as 'bumps' above the general shape of this background spectrum (Donoghue et al., 2020). Recently, Brookshire (2022) pointed out that shuffling-in-time analyses remove all temporal structure from the data, both periodic and aperiodic, producing a null spectrum with roughly equal amplitude at each frequency. This 'white noise' null distribution is likely to differ from the amplitude spectrum produced by unshuffled data simply because there is aperiodic structure in the real data, regardless of whether or not an oscillation is present. As such, it is possible that past results produced using the shuffling-in-time method reflect aperiodic structure rather than true behavioural oscillations.

Brookshire (2022) proposed two methods to quantify the presence of behavioural oscillations while controlling for aperiodic temporal structure. Here we focus on the AR-surrogate method, as this method showed superior performance over the sampling rates and trial durations that are typical of behavioural oscillations designs. In this method, the group-averaged behavioural time-course is fit with an auto-regressive model with a single coefficient – a.k.a., an AR(1) model – that is able to capture aperiodic structure in the data but unable to model additional periodic patterns. The fitted AR(1) model is used to produce a large number of simulated behavioural time-courses that are then Fourier transformed to produce a distribution of null spectra with similar aperiodic properties to the real data.

Conceptually, the AR-surrogate approach addresses the problem of aperiodic influences that might produce false positives in the detection of behavioural oscillations. Nonetheless, this method has sparked considerable debate as it has low sensitivity to detect oscillations of realistic behavioural magnitudes (Brookshire, 2023; Fiebelkorn, 2022; Re et al., 2022; Vinck et al., 2022). For example, previous behavioural oscillation studies have shown that hit-rates in detection paradigms typically fluctuate by 15–20% during the second following a reset event (for example, hit rates might fluctuate between 65% and 85%, for a fluctuation magnitude of 20%). However, the AR-surrogate method in the form proposed and tested in Brookshire (2022) does not have adequate statistical power to detect simulated behavioural oscillations below a magnitude of 30%. Statistical power was even lower for behavioural-oscillation frequencies of 6 Hz and below. Thus, while theoretically appropriate, the AR-surrogate method may not be practically useful in its current form due to its propensity to give false-negative results in the presence of effects of realistic magnitude.

The high false-negative rate of the AR-surrogate method when applied to simulated data creates a difficulty in the interpretation of null results when the same method is applied to real data. Brookshire (2022) re-analysed four publicly available datasets that had previously shown evidence of significant behavioural oscillations. Using those methods, non-significant results were produced for every test. Such a result might be interpreted as evidence that prior results using the shuffling-in-time method reflect aperiodic temporal structure rather than true behavioural oscillations. However, given the high false-negative rate for simulated effects of similar magnitude, the non-significant results observed with Brookshire's methods could alternatively reflect false-negatives.

One limitation of the Brookshire's (2022) AR-surrogate method is that the analysis is run on the final group-averaged behavioural time-course. This means that the typical approach taken in psychology when faced with a small effect, that of testing larger groups of participants, in this case will not be effective for increasing the power of the test. The approach we propose here is a simple modification of the AR-surrogate method that addresses this limitation. Instead of applying the analysis to the group-averaged time-course, we compute the same AR(1) surrogate analysis for each individual participant and then perform group-level statistical tests on the deviations of their true behavioural spectrum from their individual-level surrogate spectra. This allows us to leverage the increased

statistical power for detecting small effects that results from employing larger samples. Using simulations, we show this approach increases the statistical power of the AR-surrogate method without increasing the false positive rate at typically examined frequencies. It makes the method sensitive to behavioural oscillations of realistic magnitudes even with as few as 10 participants. It further increases the sensitivity of the analysis to behavioural oscillations in the commonly reported 3-6 Hz frequency range. Finally, we show that, when the same publicly available datasets analysed by Brookshire (2022) are re-analysed with this method, significant evidence for the existence of behavioural oscillations is obtained. Code for performing the participant-level AR-surrogate analysis in MATLAB is available at https://github.com/AnthMHarris/BehavOsc-ARSurrogate-PLevel/.

Methodological Details

This analysis is a simple modification of Brookshire's (2022) AR-surrogate approach that calculates the deviation from aperiodicity for each individual, rather than for the group average. As such, this analysis addresses a subtly different question to that being asked in analyses of the group-averaged behavioural time-course. In an analysis of the group-averaged time-course, a statistical test for periodicity is asking whether significant evidence for periodicity is present in the average of participants' behavioural reports. Such a test assumes that each individual dataset shows a similar pattern, such as an oscillation with the same frequency and phase. A test of the group-level time-course provides only indirect evidence for the presence of oscillations within the participant-level data as it is possible for patterns to emerge at the group level that are not present in any individual dataset, such as when subsets of data with different behavioural profiles are averaged. In comparison, applying the test at the individual level then computing group-level statistical tests asks whether, on average, evidence for rhythmicity is present directly in the participant-level data. It tests whether different individuals' data contain rhythmicity at the same frequency, but makes no assumptions about similarity of phase angles across individuals. As these experiments are intended to assess the presence of rhythmicity in participant-level phenomena such as perception and attention, participant-level analyses are a more direct test of the theoretically relevant information.

Below we describe the approach taken in the participant-level AR-surrogate analysis, focussed solely on the treatment of the participant-level behavioural time-courses. Prior to the steps described here, the single-trial raw data will need to be pre-processed, cleaned, detrended, transformed, aggregated, etc., in whatever manner is required by the experimental design. These steps are not covered here.

Once the behavioural time-course is obtained for a given participant, it is then fit with an AR(1) model.

$$X_t = \delta + \varphi_1 X_{t-1} + \varepsilon_t$$

In which X_t is the value of the time-series at time *t*, δ is a constant, $\varphi_1$is the autoregressive coefficient of the model that quantifies the influence of the value at time *t-1* on the value at time *t,* and ε_t is a noise term. This model is used to generate a large number of surrogate time-courses (e.g., 1000) with similar aperiodic properties to the original data. A method of spectral decomposition (here, the Discrete Fourier Transform) is then used to transform the true behavioural time-course into its spectral representation, focussing on the frequency range from 3-20 Hz (see simulations below). Next, the aperiodic surrogate spectrum is computed by applying the same spectral transformation to each of the surrogate time-courses and averaging the resulting surrogate spectra. This aperiodic surrogate spectrum reflects the aperiodic patterns present in the data but will not capture any oscillations that might be present. Thus, after subtracting the aperiodic surrogate spectrum from the true behavioural spectrum, we are left with a noisy representation of any periodic content in the data. We term these difference values the *periodic spectrum*, to distinguish them from the aperiodic spectrum that has been removed via subtraction.

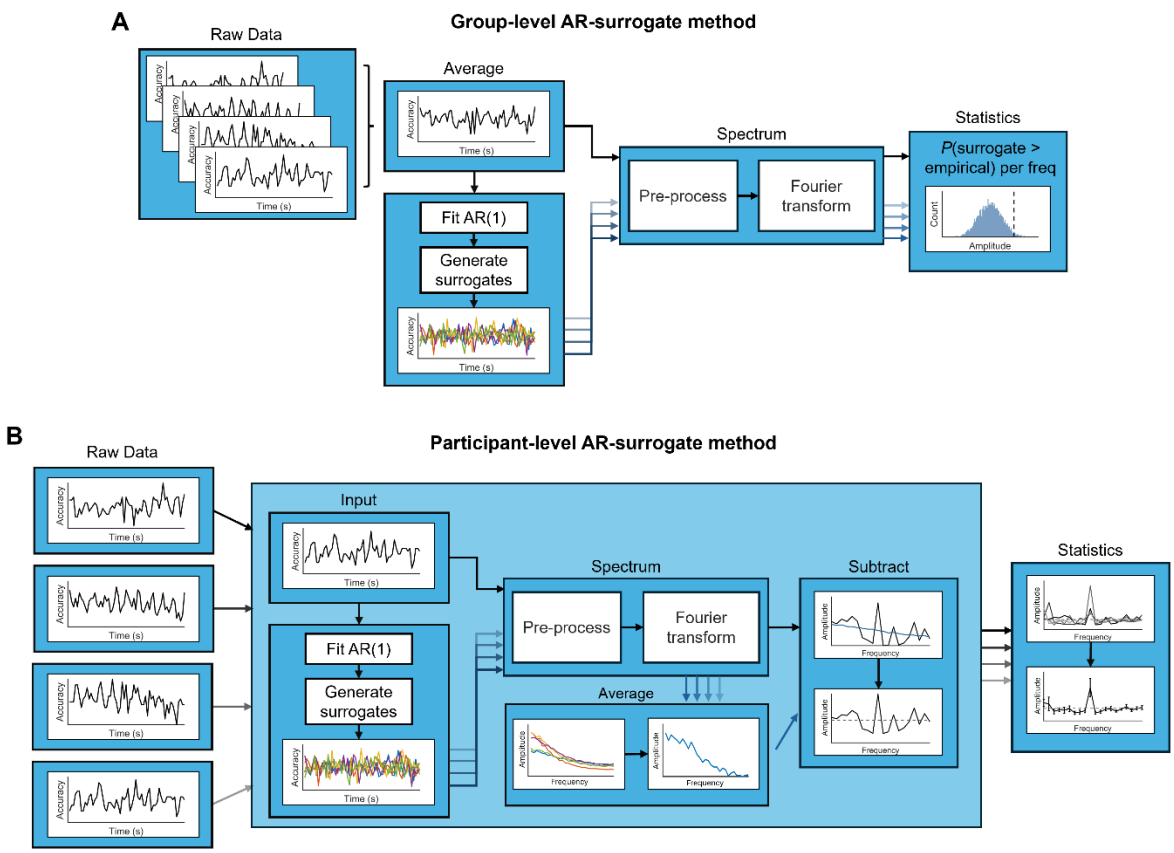

Figure 1. Differences between the group-level and participant-level AR-surrogate analyses. A) Brookshire's (2022) group-level AR-surrogate analysis. The participant-level behavioural time-courses are averaged and the group-average time-course is fit with an AR(1) model and used to generate a large number of surrogate time-courses. The real data and the surrogates are pre-processed in the same manner then Fourier transformed. The surrogate spectra are used to produce a null distribution of amplitude values at each frequency against which to compare the amplitude values of the real data. Typical corrections for multiple comparisons are applied (not shown here). **B)** For the participant-level AR-surrogate analysis, each of the participant-level time-courses are fit with an AR(1) model which is used to generate a large number of surrogate time-courses. The time-courses are pre-processed and Fourier transformed, then the average of the surrogate spectra is subtracted from the amplitude spectrum of the real data to produce a periodic spectrum. The periodic spectra for each participant are then analysed with typical statistical approaches (e.g., t-tests). Typical corrections for multiple comparisons are applied (not shown here). Panel A adapted from Brookshire (2022) under Creative Commons Attribution 4.0 International License: http://creativecommons.org/licenses/by/4.0/.

Oscillations always appear as peaks in a amplitude spectrum relative to any aperiodic background activity. Thus, only positive values in the periodic spectrum have a meaningful interpretation. At the single-participant level, the periodic spectra are extremely noisy, making it unclear which peaks reflect genuine periodicities in the data. To determine which peaks reflect behavioural oscillations at a particular frequency, we apply standard group-level statistical approaches, to capture the variability among observations of individuals and statistically assess the between-persons reliability of the oscillation presence. As we are only interested in positive values in the periodic spectrum, we can employ one-tailed one-sample *t*-tests at each frequency, then correct for multiple comparisons with one of the many available options. Here, we employ the Bonferroni-Holm procedure (Holm, 1979). Peaks that are significant at the group level after correcting for aperiodicity in the data suggest the presence of an oscillation in the participant-level behavioural time-courses.

Validation via Simulation

The intention of the first simulation was to test whether the proposed method can recover behavioural oscillations of a known frequency. We simulated participant-level data from a

behavioural oscillations task with accuracy as the metric. Each participant's data was produced by creating an independent idealised time-course of 1-second length, sampled at 60 Hz. This probability time-course was the sum of $1/f^2$ 'Brownian' noise with a 10 Hz oscillation, and was used to generate 20 random Bernoulli observations (hit = 1, miss = 0) at each time point, with the probability of a hit determined by the idealised time-course(**Figure 2A-C**). The average of observations at each timepoint produced the Response Time-course, akin to a behavioural time-course from a typical experiment (**Figure 2D**). The amplitude of the simulated oscillation corresponds to the effect size of the behavioural oscillation, which we refer to as the *behavioural depth modulation*. For example, a depth modulation of 20% indicates that the behavioural oscillation fluctuated by a difference of 20% accuracy between its peak and its trough (e.g., between 65% at the trough and 85% at the peak). We repeated this simulation at depth modulations of 0% (oscillation absent) to 50% in steps of 5%. In this simulation, the phase of the behavioural oscillation was randomly determined for each simulated participant. This procedure was performed for 500 simulated experiments at each level of depth modulation.

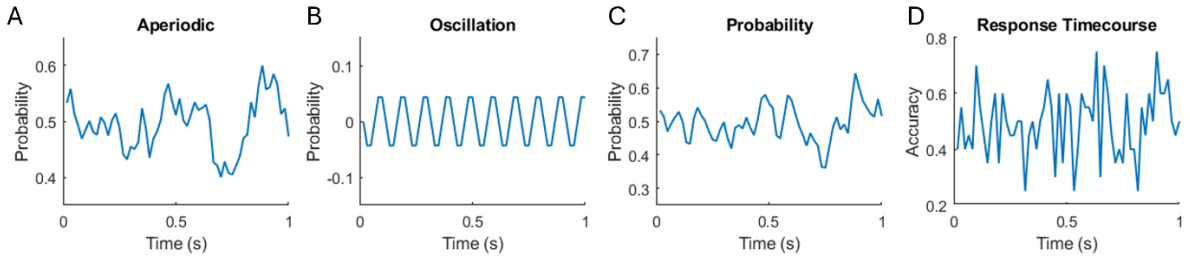

Figure 2. Example data simulation. The behavioural responses were simulated by creating a probability time-course (example shown in **C**) that was the sum of an aperiodic component (example shown in **A**) and an oscillatory component (example shown in **B**). After producing the probability time-course, the behavioural response time-course (**D**) was produced by comparing one random number for each simulated trial to the probability value at the relevant time-point for that trial. Values above the probability threshold were considered 'hits'. This example reflects the process for one participant in one simulated experiment with a behavioural depth modulation of 10%.

Following the production of the simulated time-courses, we performed the analysis described above, correcting the final p-values for multiple comparisons across frequency using the Bonferroni-Holm procedure. In each simulated experiment, a significant behavioural oscillation at 10Hz is treated as a correct detection, allowing us to calculate the statistical power of such an approach. A significant oscillation detected at oscillation magnitude of 0% (oscillation absent) indicates a false positive and is used to calculate the false-positive rate. We also examine false positives produced at frequencies other than the simulated oscillation frequency.

The example results presented in **Figure 3** are for simulated experiments with 10 participants each. The participant-level AR-surrogate method had high power to detect behavioural oscillations at the target frequency, 10 Hz, and produced very few false positives. Indeed, the false positive rate was below 5% for both the oscillation absent condition (depth modulation of 0%), and for frequencies outside 10 Hz. The following simulations are aimed at comparing the sensitivity of this method against Brookshire's (2022) group-level AR-surrogate method and testing the robustness of these findings across different numbers of participants, trials, and across different target frequencies.

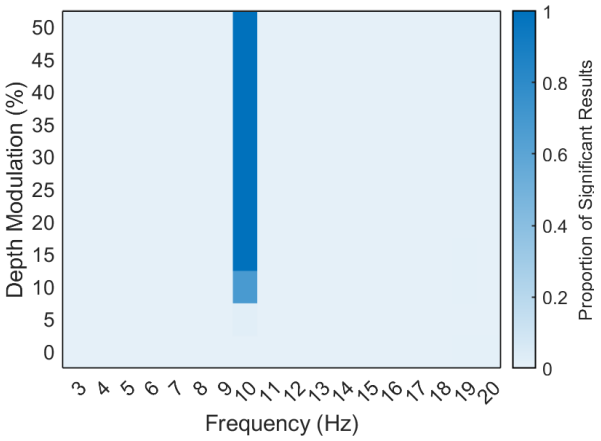

Figure 3. Example results for a 10 Hz behavioural oscillation. This simulation involved 500 experiments per depth modulation, each with 10 simulated participants. Sensitivity to detect the behavioural oscillation was above 0.95 for all depth modulations of 15% and above.

Increasing Sensitivity by Increasing Sample Size

The purpose of this simulation was to compare the sensitivity of the participant-level AR-surrogate method with the group-level AR-surrogate method proposed by Brookshire (2022), and to quantify the relationship between number of participants and sensitivity for the two methods. We followed the method described above to generate the simulated behavioural time-courses with a 10 Hz oscillation of varying magnitude and then analysed the same time-courses with the participant-level AR-surrogate method and Brookshire's (2022) group-level AR-surrogate method. As shown in **Figure 4A**, the participant-level AR-surrogate method achieved a sensitivity of 0.9 (real effects detected 90% of the time) at a depth modulation of 15% when 10 participants were simulated. With 20-150 participants, sensitivity of 0.9 was achieved at a depth modulation of 10%. This dropped to 5% when 160+ participants were analysed. In contrast, the group-level AR-surrogate method (**Figure 4B**) did not reach a sensitivity of 0.9 until the participant-level depth modulation reached 50%, and never reached a sensitivity of 1 within the ranges tested here. Note, this result is slightly different to that shown in Brookshire (2022), as the behavioural depth modulation in that paper was defined at the group level, whereas here the depth modulation is defined at the participant level. As expected, the group-level method was relatively insensitive to the number of participants in the analysis. This analysis confirms that the participant-level AR-surrogate method has superior sensitivity to detect behavioural oscillations of realistic magnitudes at all sample sizes.

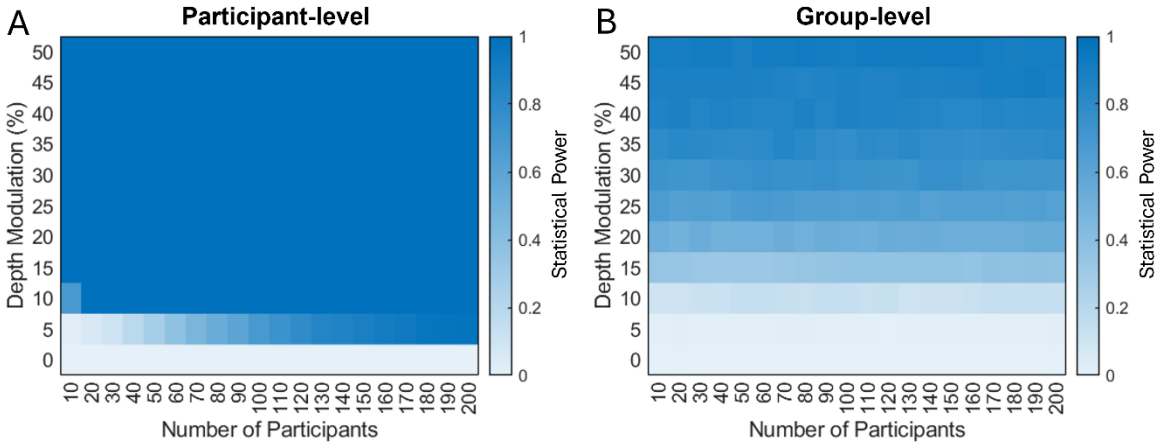

Figure 4. Comparison of participant-level and group-level AR-surrogate method. A) The results of simulated experiments (500 per cell) analysed with the participant-level AR-surrogate method. Results only shown for the target frequency (10 Hz). **B)** The same data analysed with the group-level AR-surrogate method.

Control of False Positives

To confirm that the participant-level AR-surrogate method has increased sensitivity to oscillations and not simply an increased false positive rate, we analysed the proportion of experiments in which significant results were produced at frequencies with no simulated behavioural oscillation (depth modulation of 0%). The proportion of significant results produced when there was no behavioural oscillation present was below .05 at all sample sizes, after controlling for multiple comparisons with the Bonferroni-Holm procedure (**Figure 5A**). This was also the case when examining the proportion of experiments producing significant results outside the simulated oscillation frequency of 10 Hz (**Figure 5B**). The participant-level AR-surrogate method controlled the false positive rate adequately in all cases. Indeed, false positives were below the expected level due to a weak bias against finding significant results in this analysis that will be explored below in the Frequency Limits analysis.

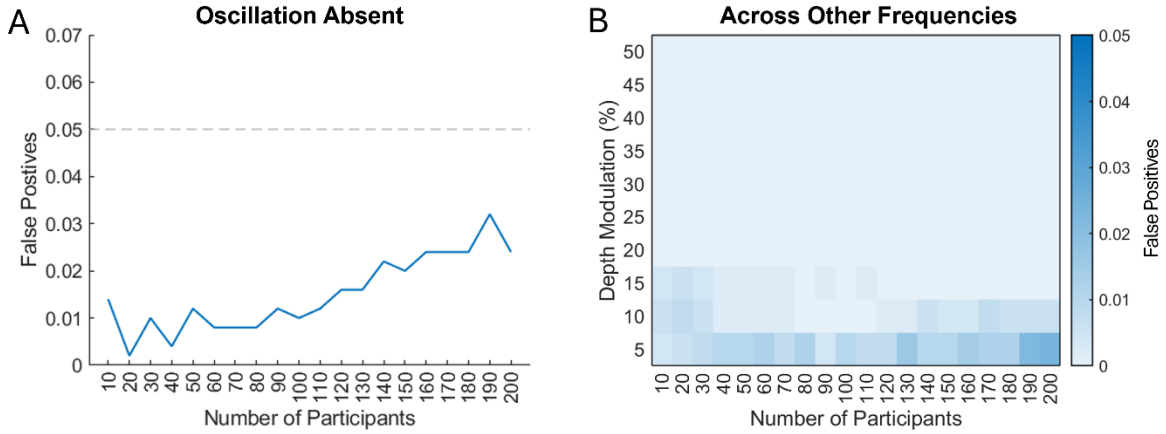

Figure 5. Proportion of false positives. A) Proportion of simulated experiments producing false positives when the behavioural oscillation was absent (Depth Modulation of 0%). For this analysis, a significant oscillation at any frequency was deemed a false positive. **B)** Proportion of simulated experiments producing significant behavioural oscillations at any frequency other than the simulated oscillation frequency.

The Effect of Trial Numbers

All the simulations up to this point have been performed with 20 trials per time bin. However, many published studies have employed fewer trials than this. To examine the influence of this factor on the sensitivity of the participant-level AR-surrogate approach, we simulated experiments with different numbers of trials per time bin. **Figure 6A** shows the sensitivity to detect 10 Hz behavioural oscillations of different depth modulations with different numbers of trials per time point, specifically for a lowest sample size simulated, 10 participants. At this sample size, a sensitivity of 0.9 for detecting behavioural oscillations was not achieved until a depth modulation of 10% with 30 trials per time point. At depth modulations of 15%, the number of trials required was reduced to 11. For a depth modulation of 20%, only 6 trials per time point were required. As expected, the number of trials per timepoint that were required to detect an effect was reduced with increased sample sizes. With 20 participants, the number of trials required to detect behavioural oscillations of 10%, 15%, and 20% depth modulations were reduced to 14, 6, and 3, respectively.

Figure 6B shows the false positive rate for different sample sizes and different numbers of trials per time-point. With 5 or more trials per time point the false positive rate was below .05 for all sample sizes. With fewer trials per time-point, more participants were required to control the false positive rate. For example, with 4 trials per time point, 50 participants were required to reach a false positive rate below .05.

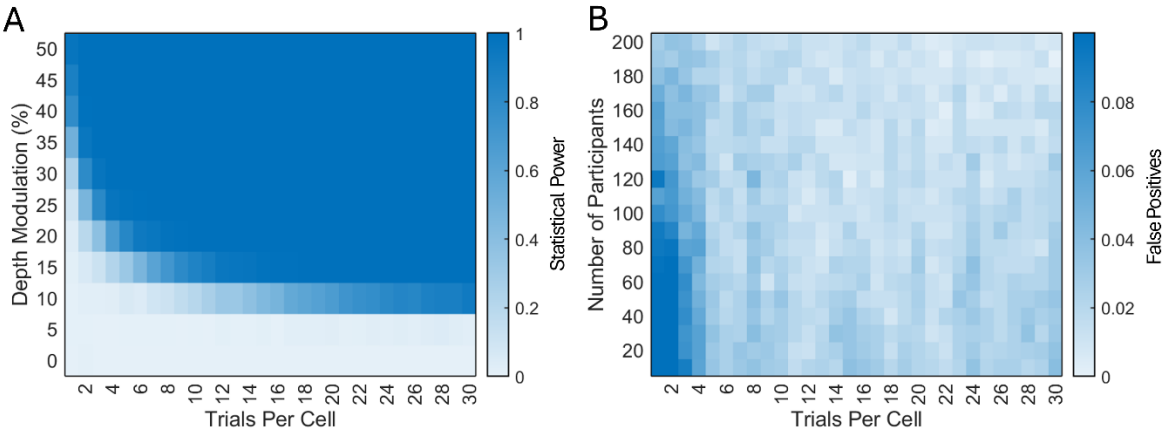

Figure 6. Effect of trial numbers. A) The sensitivity of the participant-level AR surrogate analysis increases with more trials-per-time-point and with increasing depth modulation. The sensitivity for a sample size of 10 participants is shown here. **B)** False positives demonstrated in the oscillation-absent condition (depth modulation 0%). False positives are above 5% when fewer than 4 trials per cell are employed.

Sensitivity to Oscillations of Different Frequencies

Up to this point, all the simulations have been performed with behavioural oscillations at a frequency of 10 Hz. Here, we evaluate the sensitivity of this method to detect behavioural oscillations of other relevant frequencies (with 20 trials per cell). Notably, Brookshire (2022) found with the group-level AR-surrogate method that the sensitivity to detect behavioural oscillations was reduced for oscillations at 6 Hz and below. **Figure 7A** shows the sensitivity to detect behavioural oscillations of different frequencies and depth modulations, specifically for a set size of 10 participants. As this figure shows, sensitivity is affected very little by the frequency of the behavioural oscillation, although there is a gentle trend towards greater sensitivity at higher frequencies. This trend is demonstrated more clearly in **Figure 7B**, which shows the sensitivity to detect behavioural oscillations for different frequencies and different sample sizes at a smallest behavioural depth modulation, 5%. The sensitivity to detect behavioural oscillations increased with both frequency and sample size. It should be noted, however, that this pattern was primarily present for the 5% depth modulation. For a depth modulation of 10%, sensitivity was well above .9 for all frequencies at all sample sizes of 20 participants or more. Once depth modulations reach typically observed magnitudes of 15%, sensitivity was perfect at all frequencies and sample sizes.

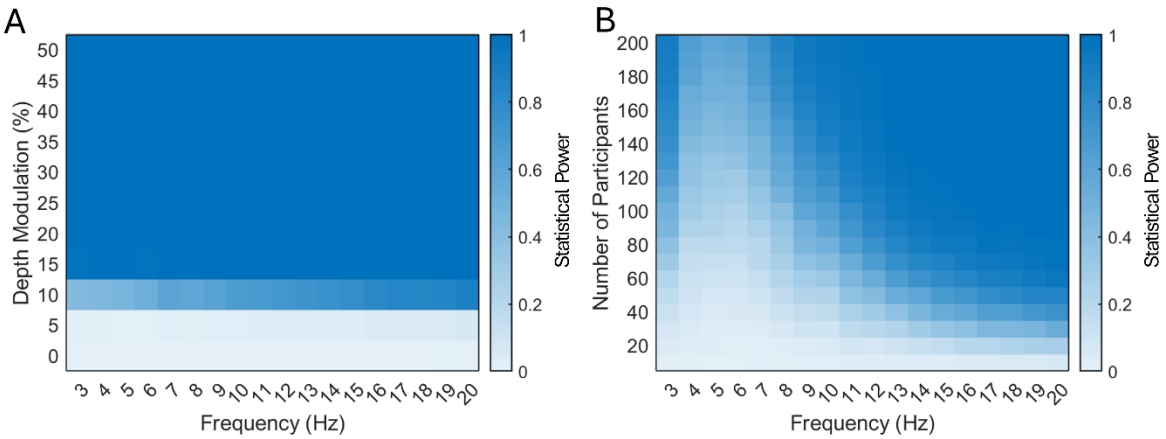

Figure 7. Sensitivity to oscillations of different frequencies. A) Target frequency by depth modulation. Example shown here is for experiments with 10 simulated participants. Sensitivity is further increased at all higher participant numbers. **B)** Target frequency by number of participants. Example shown here is for the weakest depth modulation tested, 5%. For depth modulations of 10% or more, sensitivity was above 0.9 for all frequencies for participant numbers >20.

Effect of Frequency Limits on False Positives

In the simulations presented above, the analysis has been limited to the range between 3 Hz and 20 Hz. In this simulation we examine the results outside this range. When the AR-surrogate method is used across the full range from 1 Hz to 30 Hz, false positives are produced in the range of 1-2 Hz, and above 23 Hz (**Figure 8A**). This is because the surrogates produced by the fitted AR(1) process underrepresent the exponent of the aperiodic amplitude function present in the original data (**Figure 8B**), leading to underestimates of amplitudes at the lowest and highest frequencies. This leads to systematic biases that produce false positives at the lowest and highest frequencies. It should be noted that these same biases are present in the group-level AR-surrogate method as well, however false positives were not observed in Brookshire (2022) due to the extreme width of the surrogate distribution at these frequencies. This problem arises due to the difficulty of capturing the full variability in the data when fitting an AR(1) process to only a small number of time-points (Krone et al., 2017; Wagenmakers et al., 2004). Systematic biases in the range from 3-20 Hz are toward higher spectral amplitudes in the surrogates than is truly present in the data, which weakens our sensitivity to detect effects, but does not produce false positives when one-sided tests are used. Thus, as shown above, when employed over the recommended 3-20 Hz range, the participant-level AR-surrogate method has no problem with false-positives. The vast majority of behavioural oscillations examined in the literature are within the 3-20 Hz frequency range, so these frequency limits are not likely to be a problem.

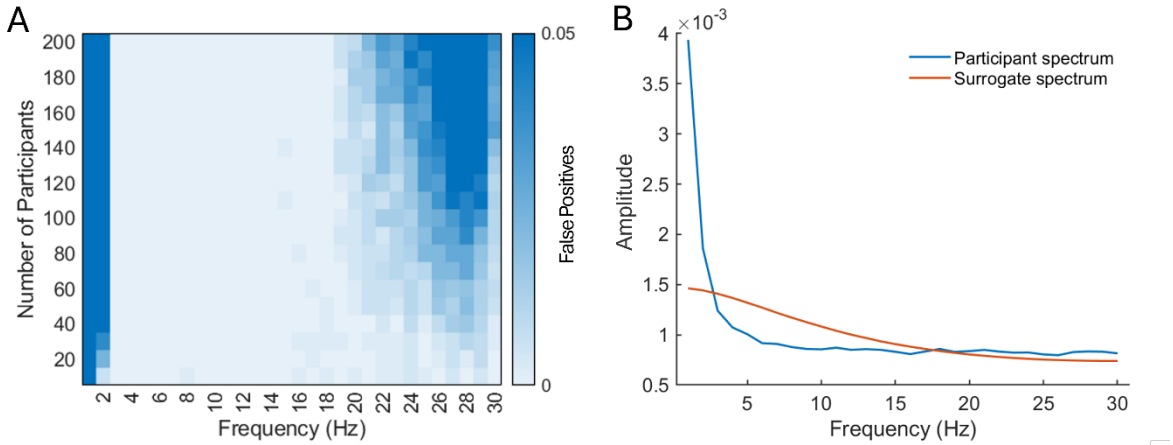

Figure 8. False positives across frequency. A) In the noise-only condition, false positives are produced outside the recommended 3-20Hz range. **B)** This is due to the AR(1) process not capturing the full variability of the data. The blue line shows the average amplitude spectrum produced from 500 simulated experiments of 10 participants each with no oscillation present. The red line shows the average amplitude spectrum from 1000 surrogate time-courses generated from the AR(1) processes fitted to each simulated participants' time-course. The surrogate spectrum underestimates amplitudes at the lowest and highest frequencies, and overestimates amplitudes in the mid-frequency range.

Reanalyses

Brookshire (2022) reanalysed 4 publicly available datasets the group-level AR-surrogate method and found no significant behavioural oscillations in any dataset. This raised questions about the existence of behavioural oscillations as a phenomenon, as their only supporting evidence came from shuffling-in-time methods that do not control for aperiodic patterns in the data. Here, we use the participant-level AR-surrogate correction method to reanalyse 3 of the 4 datasets analysed by Brookshire (2022) and find significant behavioural oscillations in all three. The 4th dataset, Davidson et al. (2018), was not analysed here because the single-participant data was not suitable for use with this method. This is because the single-participant data did not make up smooth time-courses but was populated mostly by zero-values interspersed with sharp spikes of non-zero values that only formed a

continuous time-course in the group average[1]. In each analysis, the data were detrended as in the original paper.

It is important to reemphasize at this point that the method employed here is testing a different hypothesis to tests run on the group-average time-course that have been the norm in the past (see discussion above). As such, this is not a test of whether any specific result replicates with the participant-level AR-surrogate analysis. Rather, the purpose of this reanalysis is to ask whether we find significant evidence supporting the existence of behavioural oscillations as a phenomenon in real behavioural time-courses. Consideration of the context of these results is outside the scope of the current work. Cursory details are given to aid comparison with the original manuscripts, but we do not provide complete information about the behavioural paradigms employed in these studies or possible interpretations of these results. The reader is referred to the original papers for further details.

Ho and colleagues (2017) reported 6 tests of behavioural oscillations in signal detection theory parameters of auditory detection, 5 of which were significant in the original paper. Reanalysis with the group-level AR-surrogate method produced no significant results (Brookshire, 2022). Re-examining this data with the participant-level AR-surrogate method revealed 1 significant behavioural oscillation. This was an 11 Hz oscillation in the time-course of the difference in sensitivity between the two ears, p_{bh} = .037, where p_{bh} is the Bonferroni-Holm (1979) adjusted p-value.

Senoussi and colleagues (2019) reported 8 tests of behavioural oscillations in the time-course of visual attention, 2 of which were significant after multiple-comparison correction. Again, reanalysis with the group-level AR-surrogate method produced no significant results (Brookshire, 2022). Repeating the analyses with the participant-level AR-surrogate method produced 3 significant results after correcting for multiple comparisons. These were a 5.8 Hz oscillation in the P1-P2 time-course for valid trials, p_{bh} = .019, and a 3.8Hz oscillation for the same data from invalid trials, p_{bh} = .040. Finally, there was a significant 7.7 Hz oscillation in the P1-P2 time-course for valid trials in which the probes were presented in the target hemifield, p_{bh} = .031.

Finally, Michel et al. (2021) reported 6 tests of behavioural oscillations in the time-course of the guess rate and precision parameters of mixture-modelled performance in a visual attention task. Two of these effects were significant in the original work. Reanalysis with the group-level AR-surrogate method produced no significant results (Brookshire, 2022). Repeating these tests with the current method produced 1 significant result after correction for multiple comparisons. This was a 4.8 Hz oscillation in the guess rate parameter on invalidly cued trials, p_{bh} = .003.

Discussion

There has been much debate recently regarding the existence of rhythmic components in the time-course of behavioural responses. While hundreds of articles exist demonstrating behavioural oscillations (for review, see Kienitz et al., 2022), they have been almost exclusively shown with analysis methods that cannot distinguish between periodic and aperiodic elements of the data (Brookshire, 2022). Brookshire (2022) proposed a method for assessing behavioural oscillations while controlling for aperiodic elements. This method was to fit an AR(1) model to the average group-level behavioural time-course, and use this model to generate surrogate spectra against which to compare the amplitude spectrum of the true group-level behavioural data. Brookshire (2022) showed that several past studies produced no significant behavioural oscillations when assessed with his group-level AR-surrogate method, leading to considerable debate (Brookshire, 2023; Fiebelkorn, 2022; Re et al., 2022; Vinck et al., 2022). One interpretation of Brookshire's (2022) results was the potential non-existence of behavioural oscillations as a phenomenon. That is, that all prior evidence of behavioural oscillations were in fact spurious results driven by aperiodic data trends. However, the

[1] Thanks to Matt Davidson for providing the single-participant data for their study so this could be assessed.

method had low sensitivity to detect behavioural oscillations of magnitudes commonly observed in the literature, making the interpretation of such null results ambiguous. Here, we present a modification of the AR-surrogate approach that improves the sensitivity of the method while maintaining its adequate control of aperiodic influences. Using this participant-level AR-surrogate method, we show in the same publicly available datasets that behavioural oscillations are indeed a robust phenomenon.

The adjustment we propose to Brookshire's (2022) AR-surrogate method is small: Rather than correcting for aperiodic influences in the group-average spectra, we propose they be corrected at the participant level and group-level statistical tests be applied to the resulting corrected spectra. Although the methodological change is small, its implications are profound. Namely, that the statistical power of the test can now be increased by increasing the number of participants, allowing the method to be used to target behavioural oscillations of even relatively small effect size. Additionally, the method is now robust to a number of undesirable properties that arise from examining only the group-average time-course. For example, oscillations in the group-level time-course can be produced by only one outlier participant with an unusually strong oscillation in their data, potentially producing significant results despite the pattern being absent in most participants. Alternatively, behavioural oscillations in group-level time-courses may be cancelled out if participants show behavioural oscillations with different phases. Statistical testing of the participant-level spectra addresses these problems.

A critical theoretical advantage of the participant-level AR-surrogate method is that it takes us closer to the phenomenon that we are interested in. Tests of behavioural oscillations are intended to assess oscillations in participant-level cognitive phenomena. The assessment of these phenomena at the group level requires a number of assumptions that are not required when assessing the phenomena at the participant level. Group-level methods assume that oscillations observed in the group average are present in the data of the individuals in the group, rather than emerging as an artefact of the averaging process. They also assume that each participant shows a similar phase in their behavioural oscillation, so the oscillations will come out in the average rather than cancelling out. By using participant-level analyses we can test whether participants show evidence for behavioural oscillations without these assumptions.

Here we showed, using simulated and real data, that the participant-level AR-surrogate method outperforms the group-level AR-surrogate method in its sensitivity to detect behavioral oscillations while maintaining adequate control of the false-positive rate. One result to highlight is that of the effect of number of participants on the sensitivity of the current method while maintaining adequate control of false positives. The group-level AR-surrogate method only showed sensitivity to behavioural fluctuations of about 50% (at the participant level). This sensitivity was not affected by participant numbers. The simulations presented here suggest that with as few as 5 trials per time-point, 20 participants are sufficient to provide >90% power to detect behavioural fluctuations of 20% and above. Performance was improved with additional participants such that >90% power to detect behavioural fluctuations of 15% was achieved with 30 participants, and for behavioural fluctuations of 10% magnitude with 70 participants completing 5 trials per time-point. These results were drastically improved by increasing the number of trials per participant. Thus, the participant-level AR-surrogate method is clearly sensitive to behavioural oscillations of realistic magnitudes, when assessed with realistic numbers of participants completing realistic numbers of trials.

One troubling result of Brookshire's (2022) work was the finding that publicly available data from past studies (Davidson et al., 2018; Ho et al., 2017; Michel et al., 2021; Senoussi et al., 2019) provided no evidence for behavioural oscillations once aperiodic contributions were controlled for. Some authors rightly pointed out that we have evidence of neural oscillations influencing behaviour that does not come from behavioural studies (Fiebelkorn, 2022), however this does not speak to the

existence of oscillations in behaviour or the utility of the behavioural oscillations method of inquiry. Behavioural oscillations provide a way to examine oscillatory contributions to processing that are difficult to examine with methods such as EEG because the oscillatory modulations occur after the effect of the stimulus-induced phase reset (Harris, 2023). Behavioural oscillations may be one of the only non-invasive methods for examining phasic fluctuations in higher-order sensory and cognitive processing. As such, the significant behavioural oscillations observed here by analysing the same publicly available datasets with the participant-level AR-surrogate method (Ho et al., 2017; Michel et al., 2021; Senoussi et al., 2019) are exciting for two reasons. Firstly, because they reinforce the existence of behavioural oscillations as a phenomenon that reflect true oscillations in the data, rather than inadequate control of aperiodic trends. And, secondly, because they demonstrate the utility of the participant-level AR-surrogate method in real data.

Despite these excellent results at resolving behavioural oscillations in both simulated and real data, there is room to improve the method. The surrogate spectra underestimated the exponent of the real spectra, leading to underestimated amplitudes in the tails of the spectrum and overestimated amplitudes in the centre. This problem persisted despite several different attempts to address it with different approaches (attempts not shown here: fitting the data in loglog space, Detrended Fluctuation Analysis, and various methods for correcting the AR(1) parameter for estimation over short time series; (Donoghue et al., 2020; Kendall, 1954; Marriott & Pope, 1954; Orcutt & Winokur, 1969; Peng et al., 1994)). The reality of behavioural testing with human participants means behavioural oscillations datasets typically include short time-series with low sampling rates that together produce low spectral resolution. Estimating the properties of a noisy dataset from a small sample is difficult, whether the estimation is performed using an AR(1) model or other methods (Krone et al., 2017; Thornton & Gilden, 2005; Wagenmakers et al., 2004). Thus, some bias in the final result may be unavoidable. Of course, performance in simulated data could be improved by fitting the exact function that underlies the data, but such a method would be unlikely to transfer effectively to the analysis of real data. Critically, little is known about the shape of amplitude spectra derived from behavioural time-courses. No large, comprehensive review has been performed to determine their typical spectral properties, and whether these differ between measures such as accuracy, reaction time, signal-detection parameters, etc. Despite its limitations, the AR(1) method employed here and by Brookshire (2022) is powerful for its ability to approximate many different aperiodic noise structures. The single-participant AR-surrogate method presented here is currently the most powerful method for behavioural oscillations analysis in the literature, however, future methods that more accurately approximate the underlying amplitude spectrum from short time-courses will further improve the utility of the method.

Behavioural oscillations provide critical insights into how the brain processes information across time. However, human behaviour is incredibly noisy, incorporating robust aperiodic fluctuations in addition to its periodic elements. These aperiodic fluctuations are a challenge for the assessment of behavioural oscillations. The participant-level AR-surrogate method addresses all the problems with prior shuffling-in-time methods while providing good sensitivity to detecting even small oscillatory effects. We hope it will be of use in the study of rhythmic processes in cognition.

Code Availability

Code for performing the participant-level AR-surrogate analysis is available at https://github.com/AnthMHarris/BehavOsc-ARSurrogate-PLevel/

References

Brookshire, G. (2022). Putative rhythms in attentional switching can be explained by aperiodic temporal structure. *Nature Human Behaviour*, 1–12. https://doi.org/10.1038/s41562-022-01364-0

Brookshire, G. (2023). *Distinguishing between periodicity and consistency over trials: Reply to Re, Tosato, Fries, and Landau*. https://doi.org/10.31234/osf.io/km7e9

Busch, N. A., Dubois, J., & VanRullen, R. (2009). The Phase of Ongoing EEG Oscillations Predicts Visual Perception. *The Journal of Neuroscience*, *29*(24), 7869–7876. https://doi.org/10.1523/jneurosci.0113-09.2009

Davidson, M. J., Alais, D., Boxtel, J. J. van, & Tsuchiya, N. (2018). Attention periodically samples competing stimuli during binocular rivalry. *ELife*, *7*, e40868. https://doi.org/10.7554/elife.40868

Donoghue, T., Haller, M., Peterson, E. J., Varma, P., Sebastian, P., Gao, R., Noto, T., Lara, A. H., Wallis, J. D., Knight, R. T., Shestyuk, A., & Voytek, B. (2020). Parameterizing neural power spectra into periodic and aperiodic components. *Nature Neuroscience*, *23*(12), 1655–1665. https://doi.org/10.1038/s41593-020-00744-x

Dugué, L., Marque, P., & VanRullen, R. (2011). The Phase of Ongoing Oscillations Mediates the Causal Relation between Brain Excitation and Visual Perception. *The Journal of Neuroscience*, *31*(33), 11889–11893. https://doi.org/10.1523/jneurosci.1161-11.2011

Dugué, L., Roberts, M., & Carrasco, M. (2016). Attention Reorients Periodically. *Current Biology*, *26*(12), 1595–1601. https://doi.org/10.1016/j.cub.2016.04.046

Fiebelkorn, I. C. (2022). There Is More Evidence of Rhythmic Attention than Can Be Found in Behavioral Studies: Perspective on Brookshire,. *Journal of Cognitive Neuroscience*, *35*(1), 128–134. https://doi.org/10.1162/jocn_a_01936

Fiebelkorn, I. C., Saalmann, Y. B., & Kastner, S. (2013). Rhythmic Sampling within and between Objects despite Sustained Attention at a Cued Location. *Current Biology*, *23*(24), 2553–2558. https://doi.org/10.1016/j.cub.2013.10.063

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

Fries, P. (2023). Rhythmic attentional scanning. *Neuron*, *111*(7), 954–970. https://doi.org/10.1016/j.neuron.2023.02.015

Haegens, S., Nácher, V., Luna, R., Romo, R., & Jensen, O. (2011). α-Oscillations in the monkey sensorimotor network influence discrimination performance by rhythmical inhibition of neuronal spiking. *Proceedings of the National Academy of Sciences*, *108*(48), 19377–19382. https://doi.org/10.1073/pnas.1117190108

Harris, A. M. (2023). Phase resets undermine measures of phase-dependent perception. *Trends in Cognitive Sciences*, *27*(3), 224–226. https://doi.org/10.1016/j.tics.2022.12.008

Harris, A. M., Dux, P. E., & Mattingley, J. B. (2018). Detecting Unattended Stimuli Depends on the Phase of Prestimulus Neural Oscillations. *Journal of Neuroscience*, *38*(12), 3092–3101. https://doi.org/10.1523/jneurosci.3006-17.2018

Ho, H., Leung, J., Burr, D. C., Alais, D., & Morrone, M. (2017). Auditory Sensitivity and Decision Criteria Oscillate at Different Frequencies Separately for the Two Ears. *Current Biology*. https://doi.org/10.1016/j.cub.2017.10.017

Hogendoorn, H. (2016). Voluntary saccadic eye movements ride the attentional rhythm. *Journal of Cognitive Neuroscience*, *28*(10), 1625–1635.

Holm, S. (1979). A Simple Sequentially Rejective Multiple Test Procedure. *Scandanavian Journal of Statistics*, *6*(2), 65–70.

Keitel, C., Ruzzoli, M., Dugué, L., Busch, N. A., & Benwell, C. S. Y. (2022). Rhythms in Cognition: The evidence revisited. *European Journal of Neuroscience*. https://doi.org/10.1111/ejn.15740

KENDALL, M. G. (1954). NOTE ON BIAS IN THE ESTIMATION OF AUTOCORRELATION. *Biometrika*, *41*(3–4), 403–404. https://doi.org/10.1093/biomet/41.3-4.403

Kienitz, R., Schmid, M. C., & Dugué, L. (2022). Rhythmic sampling revisited: Experimental paradigms and neural mechanisms. *European Journal of Neuroscience*, *55*(11–12), 3010–3024. https://doi.org/10.1111/ejn.15489

Krone, T., Albers, C. J., & Timmerman, M. E. (2017). A comparative simulation study of AR(1) estimators in short time series. *Quality & Quantity*, *51*(1), 1–21. https://doi.org/10.1007/s11135-015-0290-1

Landau, A., & Fries, P. (2012). Attention Samples Stimuli Rhythmically. *Current Biology*, *22*(11), 1000–1004. https://doi.org/10.1016/j.cub.2012.03.054

Marriott, F. H. C., & Pope, J. A. (1954). Bias in the Estimation of Autocorrelations. *Biometrika*, *41*(3/4), 390–402.

Mathewson, K. E., Gratton, G., Fabiani, M., Beck, D. M., & Ro, T. (2009). To See or Not to See: Prestimulus α Phase Predicts Visual Awareness. *The Journal of Neuroscience*, *29*(9), 2725–2732. https://doi.org/10.1523/jneurosci.3963-08.2009

Menétrey, M. Q., Vogelsang, L., & Herzog, M. H. (2022). A guideline for linking brain wave findings to the various aspects of discrete perception. *European Journal of Neuroscience*, *55*(11–12), 3528–3537. https://doi.org/10.1111/ejn.15349

Michel, R., Dugué, L., & Busch, N. A. (2021). Distinct contributions of alpha and theta rhythms to perceptual and attentional sampling. *European Journal of Neuroscience*. https://doi.org/10.1111/ejn.15154

Orcutt, G. H., & Winokur, H. S. (1969). First Order Autoregression: Inference, Estimation, and Prediction. *Econometrica*, *37*(1), 1. https://doi.org/10.2307/1909199

Peng, C.-K., Buldyrev, S. V., Havlin, S., Simons, M., Stanley, H. E., & Goldberger, A. L. (1994). Mosaic organization of DNA nucleotides. *Physical Review E*, *49*(2), 1685–1689. https://doi.org/10.1103/physreve.49.1685

Pomper, U., & Ansorge, U. (2021). Theta-Rhythmic Oscillation of Working Memory Performance. *Psychological Science*, *32*(11), 1801–1810. https://doi.org/10.1177/09567976211013045

Re, D., Tosato, T., Fries, P., & Landau, A. N. (2022). Perplexity about periodicity repeats perpetually: A response to Brookshire. *BioRxiv*, 2022.09.26.509017. https://doi.org/10.1101/2022.09.26.509017

Romei, V., Gross, J., & Thut, G. (2012). Sounds Reset Rhythms of Visual Cortex and Corresponding Human Visual Perception. *Current Biology*, *22*(9), 807–813. https://doi.org/10.1016/j.cub.2012.03.025

Senoussi, M., Moreland, J. C., Busch, N. A., & Dugué, L. (2019). Attention explores space periodically at the theta frequency. *Journal of Vision*, *19*(5), 22. https://doi.org/10.1167/19.5.22

Song, K., Meng, M., Chen, L., Zhou, K., & Luo, H. (2014). Behavioral Oscillations in Attention: Rhythmic α Pulses Mediated through θ Band. *The Journal of Neuroscience*, *34*(14), 4837–4844. https://doi.org/10.1523/jneurosci.4856-13.2014

Thornton, T. L., & Gilden, D. L. (2005). Provenance of correlations in psychological data. *Psychonomic Bulletin & Review*, *12*(3), 409–441. https://doi.org/10.3758/bf03193785

Tosato, T., Rohenkohl, G., Dowdall, J. R., & Fries, P. (2022). Quantifying rhythmicity in perceptual reports. *NeuroImage*, *262*, 119561. https://doi.org/10.1016/j.neuroimage.2022.119561

Vinck, M., Uran, C., & Schneider, M. (2022). *Aperiodic processes explaining rhythms in behavior: A matter of false detection or definition?* https://doi.org/10.31234/osf.io/wzvfh

Wagenmakers, E.-J., Farrell, S., & Ratcliff, R. (2004). Estimation and interpretation of 1/fα noise in human cognition. *Psychonomic Bulletin & Review*, *11*(4), 579–615. https://doi.org/10.3758/bf03196615

Williams, J. G., Harrison, W. J., Beale, H. A., Mattingley, J. B., & Harris, A. M. (2024). Effects of neural oscillation power and phase on discrimination performance in a visual tilt illusion. *Current Biology*. https://doi.org/10.1016/j.cub.2024.03.014