Irreversibility of EEG Data in Perceptual Decision Making

Master's Thesis in Brain and Cognition 23/24

Author: Jake Tear

Supervisor: Dr. Gustavo Deco

GUSTAVO RICARDO DECO - DNI x4823212c

Firmado digitalmente por GUSTAVO RICARDO DECO - DNI x4823212c
Fecha: 2024.06.27 19:32:37 +02'00'

Co-supervisor: Elvira del Agua

ELVIRA DEL AGUA BANYERES - DNI 49535164H

Firmado digitalmente por ELVIRA DEL AGUA BANYERES - DNI 49535164H
Fecha: 2024.07.01 16:37:50 +02'00'

27th of June 2024

Table of Contents

Acknowledgements

I dedicate this work to my late Papa, who supported my education in every possible way. And to my parents. I thank Elvira del Agua for her time and attention spent wrestling with code and concepts. And Gustavo Deco for allowing me to explore his groundbreaking thoughts and theories, for his guidance, and for bringing together such brilliant minds in the group. I thank Luca Bonatti for coordinating this Master's with rigor, compassion, and humor. And to my cohort for making Barcelona so enjoyable. I thank ChatGPT, of which my use resembled a form of prayer. And finally, to the dreamers of the golden neuroscience dream, those who believe that brain data is a secret journal written in a language we are yet to fully understand.

Abstract

The brain, as a dynamical system, is constantly adapting perception and behavior to meet internal and environmental demands, producing entropy as a result. This level of entropy production can be estimated by the system's irreversibility in time. Irreversibility in brain time-series, operationalized in the INSIDEOUT framework, has previously been related to asymmetric information flow between regions. This measure has shown its utility in the context of sleep, anesthesia, disorders of consciousness, and psychedelics; but as such hasn't been applied to the more momentary processes of perceptual decision making. This thesis explores the irreversibility of EEG data during a cued visual discrimination task in 28 healthy male participants across three pharmacological conditions: Placebo, Atomoxetine, and Donepezil. Our findings reveal that pre-stimulus irreversibility exhibits an inverted U-shaped relationship with performance (perceptual sensitivity), reminiscent of the Yerkes-Dodson law. During stimulus processing, global irreversibility and irreversibility in the centro-parietal region correlate linearly with performance. Pharmacologically, Atomoxetine seemed to balance global information flow, thus reducing irreversibility, one possible explanation for its overall improved performance at the group level. Irreversibility is also related but differentiable from other markers of arousal and cortical state in our data. Overall, these results further support irreversibility as a specific measure of asymmetric information flow, sensitive to changing cognitive demands in tight time-scales.

Introduction

Living systems produce entropy to self-interestedly and narrowly order a universe careening toward disorder (Ulanowicz & Hannon, 1987). This stands in contrast to systems in equilibrium. In equilibrium, there is no directionality in state changes and no entropy production (Gnesotto et al., 2018) (Supplemental Figure 1). The brain, on the other hand, is a complex dynamical system that operates far from equilibrium, at least in its healthy state (Deco et al. 2021). Indeed, empirical evidence has shown that when the brain is producing flexible behaviors described as "consciousness", there is a shift toward non-equilibrium (Carhart-Harris et al., 2014). However, more flexibility is not always beneficial depending on the context or the brain network (Safron et al., 2022).

The directionality of state changes in non-equilibrium are directly related to the arrow of time (Feng & Crooks, 2018). In one analogy, imagine a video of a cup breaking, which is easily distinguishable from the same video running in reverse. This distinguishability allows us to estimate the system to be further from equilibrium with higher entropy production (Deco et al., 2021). The process of a cup breaking is highly non-reversible, or high in irreversibility (Supplemental Figure 1). The INSIDEOUT framework, inspired by thermodynamics, measures the irreversibility of a dynamical system by the ability to distinguish between the forward and reverse time series (Deco et al., 2022).

Previously, irreversibility was found to be significantly higher in healthy controls compared to Alzheimer's disease patients, and positively correlated with cognitive functioning levels in those patients (Cruzat et al. 2023). This was found despite no significant correlation between EEG and fMRI functional connectivity values within the same sample, suggesting a more specific measure of inefficient brain dynamics lacking in entropy production (Cruzat et al. 2023). G-Guzmán, et al. (2023) also found higher irreversibility in healthy controls, followed by minimally conscious (awake but inconsistent cognitively mediated behaviors), followed by unresponsive wakefulness (awake but only reflexive behaviors) patients.

However, Kringelbach et al. (2023) found lower irreversibility in movie-watching compared to resting-state, yet higher levels in tasks compared to rest. And Shinozuka et al., (2024) found lower irreversibility in psychedelics compared to placebo. Important to note, irreversibility as an estimation of entropy production is different from

"entropy" as measured by the unpredictability of brain data (Shinozuka et al., 2024) (Supplemental Figure 2). Irreversibility instead is proposed to quantify the level of asymmetry in information flow between regions (or electrodes), or globally when averaged (Kringelbach, Perl, & Deco, 2024).

In perception, information flows bidirectionally along a processing hierarchy, which can be conceptualized as top-down weights (predictions) being continually updated by bottom-up prediction errors at each hierarchical level (Friston, 2010). This is directly related to the asymmetric information flow measured by irreversibility. For instance, global irreversibility captured the changes to the balance between top-down and bottom-up information flow seen in psychedelics, and the rank-ordering of regional irreversibility values corresponded to the anatomical hierarchy of cortical visual regions (Shinozuka et al., 2024).

While other papers have focused on how consciousness or disease states are indexed by irreversibility (Deco et al. 2022; Cruzat et al, 2023; G Guzman et al., 2023; Shinozouka et al., 2024), here we calculate irreversibility within short time windows (500 ms) in EEG data during a perceptual decision making task. The cued visual discrimination task, outlined in Nuiten, S.A et al., (2023), was performed under three drug conditions designed to manipulate baseline arousal state: Placebo (PLC), Atomoxetine (ATX), and Donepezil (DNP). In an exploratory analysis, we investigate the irreversibility of EEG data across varying task demands, drug conditions, pupil-linked arousal states, and oscillatory activity in the alpha band, allowing us to speculate on the mechanisms that generate asymmetric information flow in perception.

Methods

Recruitment, procedure, data collection, and pupil size preprocessing steps for this dataset are described in extensive detail in Nuiten, S.A et al., (2023) and Beerendonk et al. (2024) as part of Simon Van Gaal's group at the University of Amsterdam.

Participants

30 healthy male participants, aged 18-30, were recruited from an online research posting at University of Amsterdam. The study was approved by local ethics and medical committees. Participants underwent extensive screening and assessments to ensure mental and physical health. Written informed consent was obtained from all participants. Two participants withdrew after the first experimental session, leading to a dataset of N=28.

Design/Procedure

Participants performed a perceptual decision making task involving discriminating the orientation (Clockwise or Counterclockwise) of a Gabor patch stimulus hidden in noise (Supplemental Figure 3). Importantly, during an initial intake session, they completed a "staircasing" procedure which determined their difficulty level for subsequent testing sessions. The goal of this was to set stimulus opacity to a level that yielded 75% performance for that participant, allowing for better comparisons across subjects and experimental conditions.

The experimental sessions involved a slightly more complicated setup. Prior to stimulus onset, participants' **overt** attention remained on a central fixation point and was enforced with eye-tracking. The screen turned blank and the trial restarted if the participant lost fixation or blinked. Participants' **covert** attention was manipulated by an attentional cue (left or right) presented for 300 ms. The attentional cue was predictive of the stimulus location 80% of the time, allowing the separation of trials into validly/invalidly cued groups. After the attentional cue, (and a delay of 1000ms), the stimulus was then presented to either the left or right side of the screen for 200ms. Subjects then had to respond with the keyboard the direction of stimulus orientation. The cue location, stimulus location, and stimulus orientation were all balanced in half of the trials.

Participants completed three experimental sessions, one for each drug (40 mg ATX, 5 mg DNP, and PLC), counterbalanced at least 1 week apart. ATX elevates

catecholamine levels (noradrenaline and dopamine), and DNP elevates acetylcholine (Bymaster et al., 2002; Jelic & Darreh-Shori, 2010), the same neuromodulatory systems implicated in pupil-linked arousal (McGinley et al., 2015; McCormick et al., 2020). Exact administration schedules, physiological effects, and side effects are described in detail in Stijn et al., (2023). The design of the study included a 3 (*Drug Condition*) x 2 (*Attentional Cue*) manipulations on dependent variable *Performance (D')*, with *EEG* and *Pupil Size* recorded concurrently.

Materials/Measures

EEG preprocessing.

The preprocessing of the raw EEG data was done by Simon Van Gaal's group as follows:

The data was re-referenced to the average of two earlobe electrodes. A high-pass filter with a cut-off frequency of 0.01Hz was applied. Automatic detection of faulty EEG channels (via PyPrep toolbox using the RANSAC algorithm) (Appelhoff, S. et al., 2023). followed by interpolation of faulty EEG channels. Epochs were created from -2s to 2s, locked to stimulus onset at 0 ms. Independent Components Analysis (ICA) was used to remove eye-blink artifacts. Another high-pass filter (>1Hz) was used on a copy of the epoched data. ICA was fitted to the filtered data with 25 components and components that correlate to EOG electrodes were removed from the epoched data. Artifacts were further removed via the autoreject toolbox (Jas et al., 2017). A current source density (CSD) was computed to minimize volume conductance. Unless otherwise stated, these were all performed with functions from MNE-Python.

Irreversibility using the INSIDEOUT Framework.

INSIDEOUT describes the creation of a homologous time-reversed series and comparing time-shifted correlations, a process which can be described for multivariate brain signals at many time- and spatial- scales (Deco et al., 2022). The process is as follows:

Calculate all pairwise timeshifted correlations for the forward time-series.

$$FS_{\text{forward},ij}(\Delta t) = -\frac{1}{2}\log\left(1 - \langle x_i(t), x_j(t+\Delta t)\rangle^2\right)$$

Where $x_i(t)$ denotes a forward time-series at region (or electrode) i and $x_j(t+\Delta t)$ denotes the forward time-series at region j shifted by Δt time-points (tau parameter). FS becomes the matrix of these pairwise correlations for all regions.

$$FS_{\text{reversal},ij}(\Delta t) = -\frac{1}{2}\log\left(1 - \langle x_i^{(r)}(t), x_j^{(r)}(t+\Delta t)\rangle^2\right)$$

Where $x_i^{(r)}(t)$ denotes a reversed time-series at region (or electrode) i and $x_j^{(r)}(t+\Delta t)$ denotes the reversed time-series at region j shifted by Δt time-points.

From there we calculate FS_{diff} as the squared difference between the elements of the forward and reversed matrices.

$$FS_{\text{diff},ij} = \left(FS_{\text{forward},ij}(T) - FS_{\text{reversal},ij}(T)\right)^2$$

For *global irreversibility* at the trial level, we took the mean of this difference matrix to get one "global" value.

For *regional irreversibility* values, we found each electrode's contribution to this difference matrix by averaging their pairwise values. We plotted these as topographic plots using python's MNE toolbox (Gramfort et al., 2013).

A tau parameter of 5 timepoints (~30ms) was selected, as this value corresponded to the first local minimum of the average autocorrelation function of all trials in our dataset.

We performed INSIDEOUT on two epochs locked to stimulus onset (at 0ms), Pre-stimulus (-500 ms to 0ms), and Post-stimulus (0ms to 500 ms).

After visual inspection of the EEG time series for trials with the highest outliers of global irreversibility values, we found that outliers corresponded with high amplitude electrode drifts not caught in the EEG preprocessing steps. We corrected this by removing trials above a global irreversibility value of 0.075 for our analysis of both pre- and post-stimulus epochs.

Most analysis and visualization was completed on log transformed global irreversibility values due to extreme right skewness in raw values.

Alpha Power (8-12 Hz).

The power spectral density (PSD) of each EEG epoch (-500ms-0ms pre-stimulus and 0ms-500ms post-stimulus) was computed using Welch's method (Welch P, 1967). The absolute power was then calculated by integrating the area under the PSD curve from

8 Hz to 12 Hz using Simpson's rule. Upon a visual inspection of outliers, several trials still contained artifacts after preprocessing that we filtered out by removing the top 0.1% of alpha power trials. Log transform of alpha power was often used due to the skewness of the data (Smulders, F.T.Y. et al., 2018)

Pre-stimulus Pupil Sizes.

Average pupil sizes for the pre-stimulus epoch (-500ms to 0ms), one value per trial, were provided by Simon Van Gaal's group. Preprocessing steps are described in Beerendonk et al. (2023).

Binning Procedure.

We adapted the binning procedure used for this dataset in (Beerendonk et. al, 2023), where we first assigned each trial to one of 5 (or 10) equally populated bins. To assign bins, we grouped within subjects, drug conditions, experimental sessions, and trial run blocks and then averaged the metric in which we are comparing (pupil size, irreversibility, alpha power, etc). We created these bins separately for analysis in the pre-stimulus baseline window (-500ms-0ms) and post-stimulus processing window (0-500ms).

Performance (D').

To calculate our D' measure of performance (perceptual sensitivity) from Signal Detection Theory (Green & Swets, 1966), we found the difference between the z-scores of the hit rate and the false alarm rate for trials in that bin.

Model fits on the bins.

We separately fit linear and quadratic regression models on the bins and the dependent variable of choice (d', alpha power, etc). The regression coefficients of each model (linear or quadratic term) for each subject and drug condition were assessed via one-sample t-tests whether they differed significantly from zero. A significant p-value indicates that the corresponding term meaningfully contributes to explaining the variability in the dependent variable. We then plotted a scatterplot of the means of the bins with the means and standard error bars of the dependent variable within each bin. We only visualized significant model fits, choosing the more significant model fit to

plot the linear or quadratic trend line and standard error of the mean, shaded in for each drug condition.

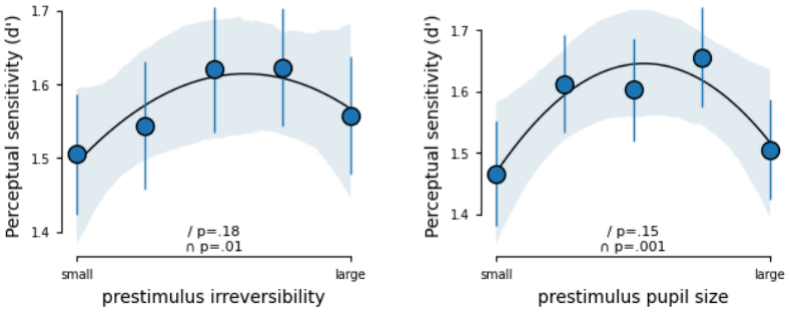
Results

Of the 44,790 trials from the 28 subjects in the full dataset (14658 in PLC; 14974 in DNP; 15158 in ATX), 44676 trials (99.7%) were included in analysis after filtering outlier alpha power and irreversibility values (14623 in PLC; 14939 in DNP; 15114 in ATX). We investigate the irreversibility of EEG data for two epochs, *pre-stimulus (-500ms to 0ms)* and *post-stimulus (0ms to 500ms)* corresponding to stimulus processing. We also relate these measures with other markers of brain state, pupil size and alpha power.

1) Pre-stimulus Irreversibility

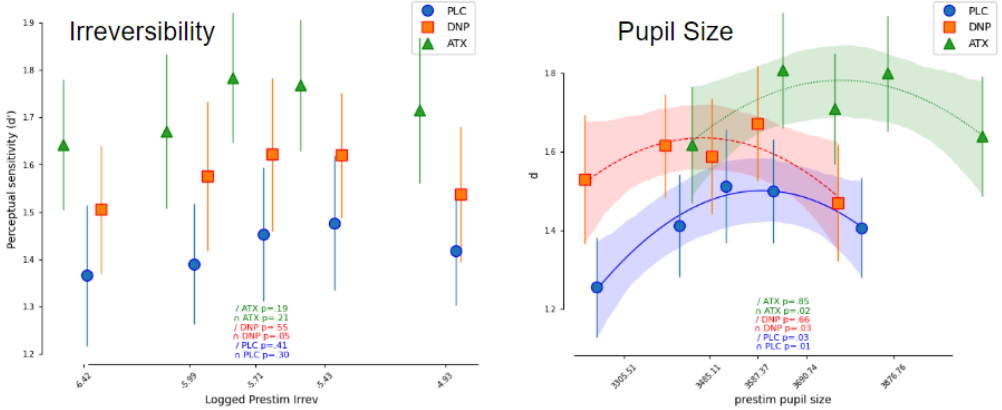
First we assessed how prestimulus global irreversibility values relate to performance (d'). The collapsed model fit for all trials is strongly significant for the quadratic inverted U relationship ($p < 0.01$). We visualized the relationship between irreversibility and d' (left) and the relationship between pupil size (arousal) and d' (right) (all trials collapsed) (**Figure 1**).

Figure 1



By drug, the inverted U relationship is close to significance in DNP (orange) ($p = 0.0505$), and similar visually in placebo (blue) and ATX (green) (PLC: $p = 0.3$; ATX: $p = 0.21$) (**Figure 2**).

Figure 2

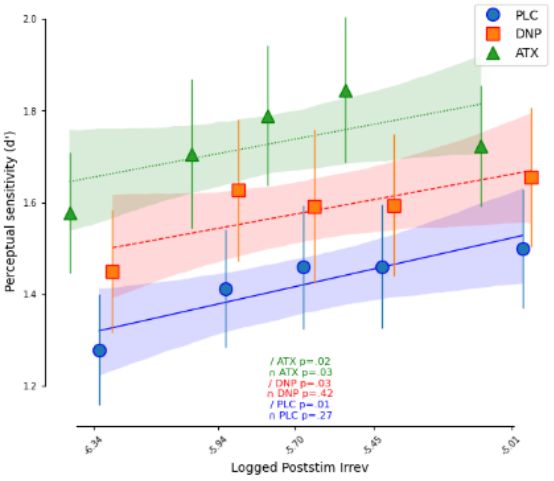


This finding supports the notion of an optimal middle level of baseline global irreversibility associated with peak performance, analogous to the Yerkes-Dodson Law for prestimulus pupil-linked arousal (Yerkes & Dodson, 1908).

2) Post-stimulus Irreversibility

In contrast to the inverted U relationship between irreversibility and performance in the pre-stimulus epoch (-500ms to 0ms), we found a significant **linear** relationship of irreversibility to performance during in the post-stimulus epoch (0ms to 500ms) for all individual drug conditions (PLC: $p < 0.05$; DNP: $p < 0.05$; ATX: $p < 0.05$) (**Figure 3**) and collapsed for all trials ($p < 0.05$).

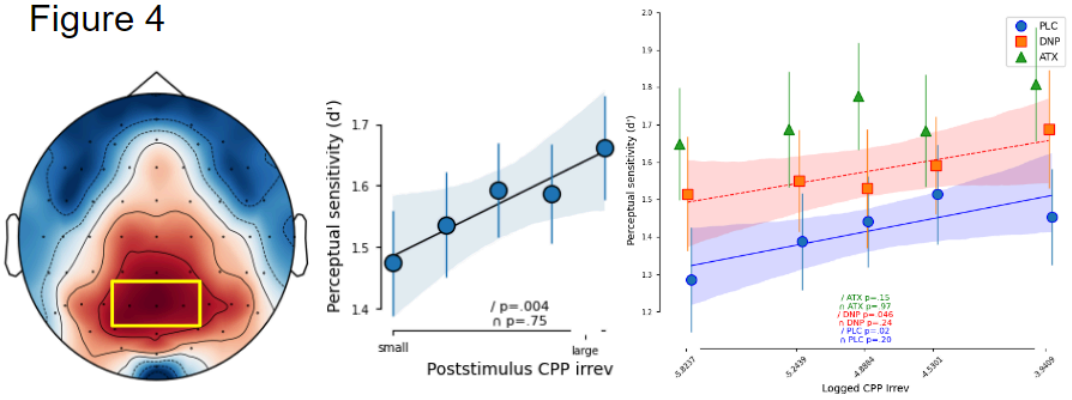
Figure 3



3) Centro-Parietal Irreversibility

We also investigated regional irreversibility using the individual EEG electrode's contribution to the reference matrix (see methods). We looked at post-stimulus regional irreversibility in electrodes Cpz, Cp1, and Cp2. Similar to the post-stimulus global irreversibility values, we found a linear relationship in the average irreversibility of these electrodes to performance (d'), which was significant for all trials ($p < 0.05$). And by drug, significant in PLC and DNP ($p < 0.05$), and trending for ATX ($p = 0.15$) (**Figure 4**).

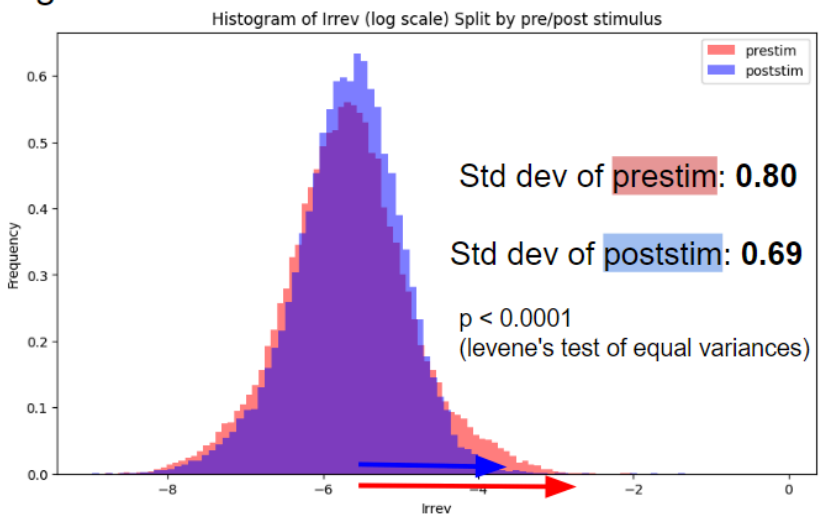
Figure 4



4) Pre-stimulus to Post-stimulus change in variability

We found a significant reduction in the variability from pre-stimulus to post-stimulus global irreversibility values for all trials and drug conditions. In all trials collapsed in the raw data, the standard deviation of pre-stimulus was 0.00478, and in post-stimulus: 0.00321 (Levene's test of equal variances, test statistic = 656.6, p-value < 0.0001). And all trials collapsed in the logged data, standard deviation of pre-stimulus: 0.80, post-stimulus: 0.69 (Levene's test of equal variances, test statistic in raw data = 666.5, p-value < 0.0001) (**Figure 5**). This relationship between pre-stimulus and post-stimulus variability is the same and highly significant for all drug conditions.

Figure 5



5) Drug effects on Irreversibility

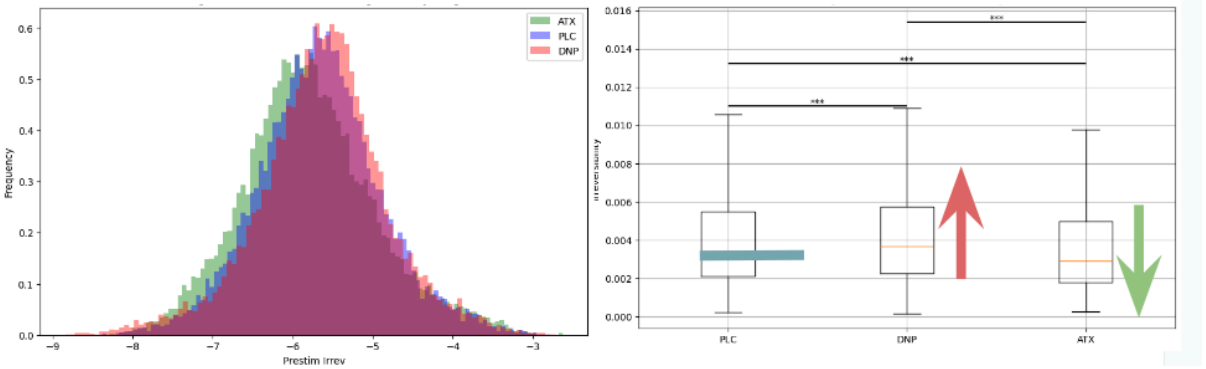
Finally, we investigate the difference in global irreversibility between the drug conditions. In the pre-stimulus epoch we report the lowest irreversibility in ATX (mean: 0.00439), followed by PLC (mean: 0.00468), with DNP as the highest (mean: 0.00491) (**Figure 6**). All pairwise comparisons were highly significant using Wilcoxon rank-sum tests, corrected for multiple comparisons ($p < 0.0001$), albeit with low effect sizes (however these effect sizes may be biased lower due to logged irreversibility data being not exactly normally distributed) (Cohen, 1988). The effect size for each pairwise comparison in the logged data: PLC vs. ATX, Cohen's $d = 0.16$; PLC vs. DNP, Cohen's $d = 0.05$; ATX vs. DNP, Cohen's $d = 0.21$.

This relationship between the drug conditions remained the same in post-stimulus (ATX mean: 0.00397, PLC mean: 0.00427, DNP mean: 0.00445) and highly significant in the post-stimulus epoch for all pairwise comparisons ($p < 0.0001$).

Effect sizes: PLC vs. ATX, Cohen's d = 0.17; PLC vs. DNP, Cohen's d = 0.06; ATX vs. DNP, Cohen's d = 0.22.

Behaviorally, Stijn et al., (2023) reported also a significant pairwise improvement in perceptual sensitivity (d') and pupil-linked arousal for ATX vs. PLC in this dataset. There was no difference behaviorally or in pupil size between PLC and DNP.

Figure 6 Irreversibility by Drug



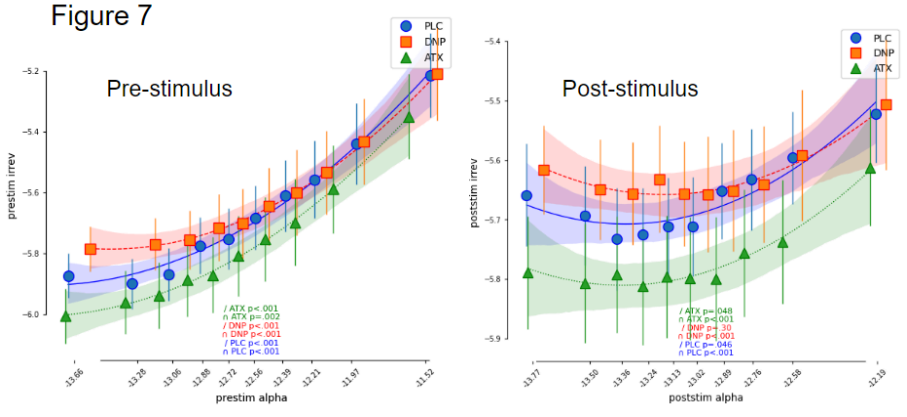
6) Irreversibility and other markers of brain state

To compare irreversibility, pupil size, and alpha power, we used the previous binning procedure, but with a total of 10 bins to more clearly show their relationships.

A) Alpha power bins and irreversibility

Interestingly, the relationship between alpha power and irreversibility was not static, the slope of the relationship clearly changes from pre-stimulus to post-stimulus (**Figure 7**). In pre-stimulus, the slope is clearly much steeper (PLC: 0.069; DNP: 0.055; ATX: 0.062) and the fit is significant in the linear model (linear model fit: $p<0.001$ for all drug conditions), that is to say, as alpha power increases, irreversibility increases. In post-stimulus, the slope flattens (PLC: 0.017; DNP: 0.010; ATX: 0.015), loses significance in the linear model fit, and becomes a positive quadratic (quadratic model fit: $p<0.001$ for all drug conditions), where at low alpha power, there are mid levels of irreversibility, middle values of alpha have the lowest levels of irreversibility, with the highest values of alpha corresponding to the highest levels of irreversibility.

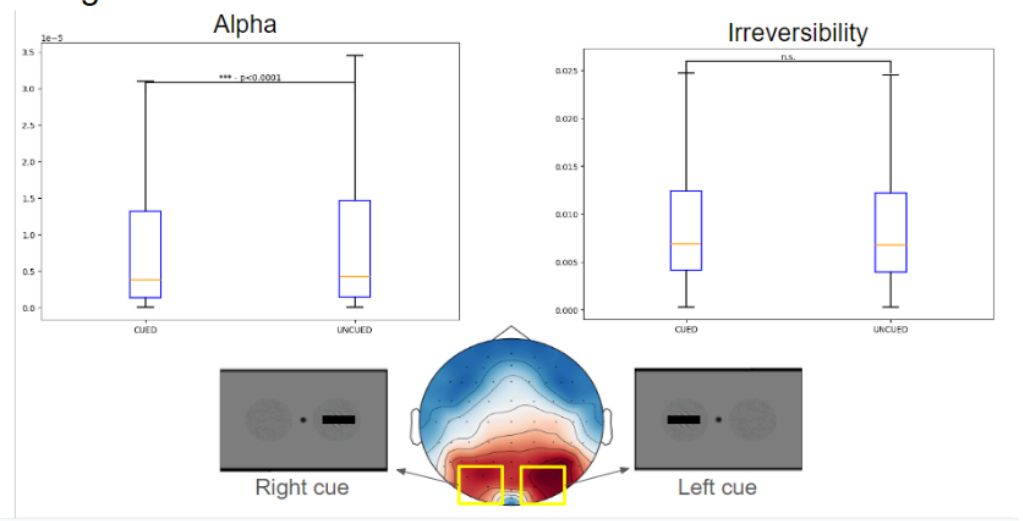
Figure 7



B) Alpha power and irreversibility in electrodes representing the attentional cue

We ran an additional analysis based on previous literature, where alpha power has been shown to decrease in cortical areas representing the location of an attentional cue (Samaha et al., 2020). We used electrodes O1, PO3, and PO7 for left hemisphere electrodes representing a rightward cue, and O2, PO4, and PO8 representing a leftward cue. Indeed we see that alpha power is lower in the cued region (mean: 1.98 x 10^{-5}, median: 4.20 x 10^{-6}) vs. uncued region (mean: 2.09 x 10^{-5}, median: 4.53 x 10^{-6}) (Wilcoxon rank sum test statistic: 2.43x10^{8}, $p < 0.0001$, Cohen's d (logged data) = 0.04). In the same time period and same electrodes, global irreversibility shows no difference (**Figure 8**).

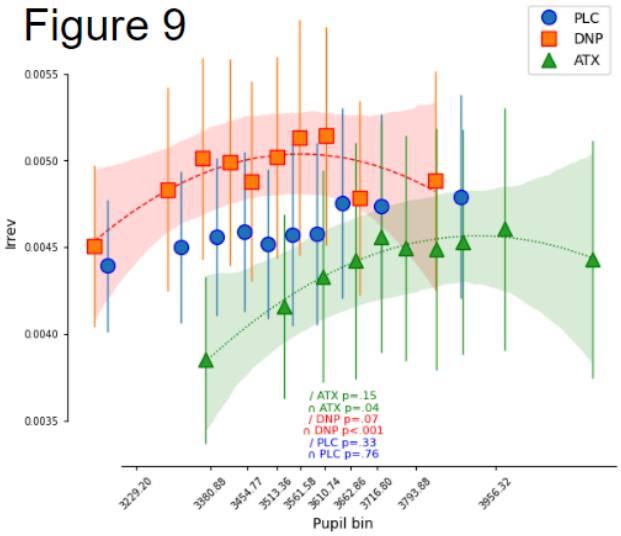
Figure 8



C) Pupil size

The relationship between prestimulus pupil size and irreversibility **(Figure 9)** appears to be drug dependent. For instance, ATX shows a trending linear ($p = 0.15$) and positively sloped inverted U relationship ($p < 0.05$). DNP shows an inverted U relationship ($p < 0.001$). And PLC's relationship appears linear but is not significant ($p = 0.33$).

Figure 9



Discussion

Overall, global asymmetric information flow as indexed by irreversibility is related to visual perception in a manner sensitive to task-demands and neuromodulation by catecholaminergic and acetylcholinergic drugs. Irreversibility is also related but dissociable from other measures of brain state like pupil-linked arousal and alpha power.

These exploratory results come with a few caveats. First, the exact relationship between incoming, outgoing, top-down, bottom-up, overall information flow, and irreversibility values are not so clear. Our measure of regional irreversibility does not distinguish between asymmetric information flow as a result of higher outgoing information or higher incoming information, or top-down vs. bottom-up information flow. Nor does global irreversibility distinguish between low information flow states and high information flow states given equal *balance* between incoming and outgoing flow across all regions. To estimate these relationships we must complete further analysis or compare our results to theory and studies investigating information flow and directionality.

1) Pre-stimulus irreversibility has an inverted U shaped relationship to performance

We found an inverted U relationship, where mid levels of prestimulus irreversibility have the highest levels of performance, with lower performance at low or high levels. We relate this to prestimulus pupil size, where in our data and other studies, there was also an inverted U relationship between prestimulus pupil size and perceptual sensitivity (d') (Beerendonk et. al, 2023; Podvalny et al., 2021). This follows the Yerkes Dodson law which states that for difficult tasks, performance is best at middle levels of arousal and lower at low or high arousal (Yerkes & Dodson, 1908). This was found despite no significant linear correlation between pupil size and irreversibility. We speculate that excess asymmetric information flow indexes a form of distraction just before the stimulus.

2) Higher irreversibility during stimulus processing improves performance

We found that post-stimulus irreversibility had a linear relationship to performance. This result is aligned with the theoretical notion that higher global irreversibility

reflects a state with higher computational demand (Kringelbach, Perl, & Deco, 2024). In other words, higher global asymmetric information flow during stimulus processing improves performance.

3) Higher irreversibility in centro-parietal electrodes improves performance

Here we looked at the centroparietal regions' contribution to irreversibility, where we took the mean irreversibility of electrodes CP1, CP2, and CPz. This regional irreversibility also showed a linear relationship to performance. In other words, more asymmetric information flow in this region is beneficial for performance. However, this was trending but not significant in ATX, possibly for reasons explained in Discussion Section 5.

In the ERP literature, a positivity 300-600ms post-stimulus in this region (the P300) is described as a neural marker of evidence accumulation leading to a perceptual decision (Twomey et al., 2015; van Vugt et al., 2019). Therefore we can speculate that the asymmetry in this region is driven by incoming information flow.

4) The variability of irreversibility is sensitive to changing task demands.

Chén, O.Y. et al., (2019) makes the case that variability in information flow patterns can serve as neural markers to predict individual behaviors. We see that the variability of irreversibility reduces significantly from the pre-stimulus to post-stimulus epoch. As far as task demands, pre-stimulus requires maintaining the representation of the attentional cue while maintaining flexibility (the cue is only right 80% of the time). It could be that this necessary behavioral flexibility is reflected in this variability measure. This matches findings from Cruzat el al. (2023), where the standard deviation of irreversibility was also higher in healthy controls compared to Alzheimer's disease patients.

Another possible explanation is that the circuits and regions that drive asymmetric information flow in the prestimulus epoch are functionally more variable than those in post-stimulus. In support of this, we do see a shift in the spatial layout of irreversibility from pre to post (Supplemental Figure 4). This means that there are slightly different regions driving the asymmetry.

Duan et al. (2023) also showed that an attention network showed less total information flow after an orienting cue. Functionally the network became more of a small world network linked by more-sparse connections. Could we be seeing the

result of a more small world network with lower (and less variable) information flow when stimulus processing is occurring?

5) Atomoxetine balances information flow, leading to lower irreversibility.

We'll focus on ATX, because DNP had no behavior effects in our data and baseline elevation of ACh doesn't appear to align with phasic and localized cholinergic signals involved in perceptual decision making (Sarter, M., & Lustig, 2020).

We propose that the lower irreversibility and higher performance associated with ATX vs. PLC is due to a balancing of information flow, however from a different route than psychedelics.

Generally, in normal resting state, there is more top-down information flow from high-order cognitive systems to primary functional systems (Chén, O.Y. et al., 2019). This can help explain why lower irreversibility was found in psychedelics vs. placebo (Shinozuka et al., 2024). According to the RElaxed Beliefs Under pSychedelics (REBUS) model, bottom-up information flow is normally shunted by top-down predictive models which "compress" the information which would otherwise be transmitted from bottom-up (related to prediction error), and that psychedelics reduce the influence of these restrictive models (Carhart-Harris & Friston, 2019; Alamia et al., 2020). The lower irreversibility seen in psychedelics therefore come from more *balanced* and less asymmetric information flow between top-down and bottom-up processes (Shinozouka et al., 2024). In line with the REBUS model, psychedelics also decrease alpha power, which seem to indicate top-down processes (Carhart-Harris et al., 2016; Benedek et al., 2011).

In a study looking at the neural dynamics of ATX, Pfeffer et al., (2021) found that, during a visual task and not rest, ATX increased cortex-wide autocorrelations, reflecting increased lateral and feedback signaling (top-down). Our task data indicate that occipital and parietal regions drive much of the asymmetry in information flow (Supplemental Figure 4), likely reflecting bottom-up signaling. The increased lateral and feedback (top-down) signaling may counterbalance the bottom-up signaling driving the asymmetry in our task data, thus reducing irreversibility. This is because top-down signaling doesn't *always* reduce bottom-up information flow like in the psychedelics example, it can also have a neutral or positive effect depending on the context (Gazzaley, A et al., 2005). And because this neural effect in ATX is context

dependent, it's possible the relationship between drug condition and irreversibility would change in resting-state data.

6) Irreversibility is related but dissociable from alpha power and pupil size.

A) Relationship between alpha and irreversibility changes to task demands

The relationship between alpha power and irreversibility is more strongly positive in prestimulus than poststimulus, and doesn't seem to be affected by drug condition. This suggests that the relationship changes based on task demands. In other words, during pre-stimulus ready-ing, the higher the alpha power, the higher the asymmetric information flow. During stimulus processing, there is a positive quadratic relationship where at low alpha power, there are middle levels of asymmetric information flow, followed by a dip in asymmetric flow at mid levels of alpha, then higher asymmetry at high levels of alpha.

This is difficult to interpret. There are mixed results in how prestimulus alpha oscillations predict perceptual performance (Waschke et al., 2019; Linkenkaer-Hansen et al., 2004; Hanslmayr et al., 2007; Iemi et al., 2017).

In our simulations, phase aligned alpha oscillations show zero irreversibility, and oscillations at linearly increasing phases show high irreversibility, indicating directionality and asymmetry in information flow (Supplemental Figure 5). It seems that alpha power and phase are important.

One caveat on our interpretation of alpha power is that direct changes to alpha power can occur as a result of broadband shifts or changes to the aperiodic component of the signal, which don't reflect genuine differences in alpha oscillations (Donoghue et al., 2020).

B) Dissociability of cortical excitability and asymmetric information flow.

When attention is experimentally manipulated with a spatial cue, alpha power typically decreases in cortical areas representing the relevant location (the contralateral hemisphere to cue) (Samaha et al., 2020). Using an example from our experimental setup, for a visual stimulus presented rightward, we'd expect (and did find) lower alpha power in the left hemisphere. We see a dissociation of the irreversibility and alpha power here, suggesting that even though alpha is active in

playing a role in changing cortical excitability, net asymmetry in information flow as indexed by irreversibility is no different in that same region and epoch.

C) Pupil size

The relationship between pupil size and irreversibility was inconclusive in our data and appeared to change by drug. Generally, alpha power and pupil size were found to be linearly coupled during quiet wakefulness, albeit with a short ~300ms time delay for pupil size (Montefusco-Siegmund et al., 2022). However, the relationship between pupil size, alpha power, and their mechanistic influences during perceptual decision making, is still not clear (Podvalny et al., 2021; Pilipenko and Samaha, 2024).

Future Directions

Some of the limitations previously mentioned can be addressed with other data analysis methods on the same data or the same methods with new task parameters. To further assess directionality of information flow, we could measure cortical traveling waves, use tests of temporal dependence between regions or networks, or compute Generative Effective Connectivity by fitting a whole-brain model on empirical irreversibility values (Alamia et al., 2020; Goddard et al., 2022; Kringelbach et al. 2023). Elshafei et al., (2022) also designed a task to evaluate both top-down and bottom-up attentional influences and found top-down specific deficits in aging. It's possible we could use a task that differentially activates top-down and bottom-up circuits to probe for the drivers of asymmetric information flow.

In conclusion, this thesis presents further evidence of irreversibility as a promising measure of asymmetric information flow, even in short time windows during perceptual decision making.

References

Appelhoff, S., Hurst, A. J., Lawrence, A., Li, A., Mantilla Ramos, Y. J., O'Reilly, C., Xiang, L., Dancker, J., Scheltienne, M., & Bialas, O. (2023). PyPREP: A Python implementation of the preprocessing pipeline (PREP) for EEG data. (0.4.3). Zenodo. https://doi.org/10.5281/zenodo.10047462

Alamia, A., Timmermann, C., Nutt, D. J., VanRullen, R., & Carhart-Harris, R. L. (2020). DMT alters cortical travelling waves. *eLife*, *9*, e59784. https://doi.org/10.7554/eLife.59784

Beerendonk, L., Mejías, J. F., Nuiten, S. A., de Gee, J. W., Fahrenfort, J. J., & van Gaal, S. (2024). A disinhibitory circuit mechanism explains a general principle of peak performance during mid-level arousal. *Proceedings of the National Academy of Sciences*, *121*(5), e2312898121. https://doi.org/10.1073/pnas.2312898121

Benedek, M., Bergner, S., Konen, ¨ T., Fink, A., & Neubauer, A.C. (2011). EEG alpha synchronization is related to top-down processing in convergent and divergent thinking. *Neuropsychologia* 49 (12), 3505–3511. https://doi.org/10.1016/j.neuropsychologia.2011.09.004

Benwell CSY, Coldea A, Harvey M, Thut G. (2022). Low pre-stimulus EEG alpha power amplifies visual awareness but not visual sensitivity. *Eur J Neurosci* 55:3125–3140. https://doi.org/10.1111/ejn.15166

Bymaster, F. P., Katner, J. S., Nelson, D. L., Hemrick-Luecke, S. K., Threlkeld, P. G., Heiligenstein, J. H., Morin, S. M., Gehlert, D. R., & Perry, K. W. (2002). Atomoxetine increases extracellular levels of norepinephrine and dopamine in prefrontal cortex of rat: a potential mechanism for efficacy in attention deficit/hyperactivity disorder. *Neuropsychopharmacology: official publication of the American College of Neuropsychopharmacology*, *27*(5), 699–711. https://doi.org/10.1016/S0893-133X(02)00346-9

Carhart-Harris, R.L., Leech, R., Hellyer, P.J., Shanahan, M., Feilding, A., Tagliazucchi, E., Chialvo, D.R., Nutt, D. (2014). The entropic brain: a theory of conscious states informed by neuroimaging research with psychedelic drugs. *Front. Hum. Neurosci.* 8, 20. https://doi.org/10.3389/fnhum.2014.00020.

Carhart-Harris, R. L., Muthukumaraswamy, S., Roseman, L., Kaelen, M., Droog, W., Murphy, K., Tagliazucchi, E., Schenberg, E. E., Nest, T., Orban, C., Leech, R., Williams, L. T., Williams, T. M., Bolstridge, M., Sessa, B., McGonigle, J., Sereno, M. I., Nichols, D., Hellyer, P. J., Hobden, P., … Nutt, D. J. (2016). Neural correlates of the LSD experience revealed by multimodal neuroimaging. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(17), 4853–4858. https://doi.org/10.1073/pnas.1518377113

Carhart-Harris, R. L., & Friston, K. J. (2019). REBUS and the Anarchic Brain: Toward a Unified Model of the Brain Action of Psychedelics. *Pharmacological reviews*, *71*(3), 316–344. https://doi.org/10.1124/pr.118.017160

Chén, O. Y., Cao, H., Reinen, J. M., Qian, T., Gou, J., Phan, H., De Vos, M., & Cannon, T. D. (2019). Resting-state brain information flow predicts cognitive flexibility in humans. *Scientific Reports*, 9(1), 3879. https://doi.org/10.1038/s41598-019-40345-8

Cruzat, J., R. Herzog, P. Prado, Y. Sanz-Perl, R. Gonzalez-Gomez, S. Moguilner, M.L. Kringelbach, G. Deco, E. Tagliazucchi, A. Ibañez. (2023). Temporal irreversibility of large-scale brain dynamics in Alzheimer's disease, *J. Neurosci*. 43, 1643–1656. https://doi.org/10.1523/jneurosci.1312-22.2022

Deco G, Sanz-Perl Y, Sitt JD, Tagliazucchi E, Kringelbach ML. (2021) Deep learning the arrow of time in brain activity: characterising brain-environment behavioural interactions in health and disease. *bioRxiv*. https://doi.org/10.1101/2021.07.02.450899

Deco G, Sanz-Perl Y, Bocaccio H, Tagliazucchi E, Kringelbach ML. (2022). The INSIDEOUT framework provides precise signatures of the balance of intrinsic and extrinsic dynamics in brain states. *Commun. Biol.* 5, 572. https://doi.org/10.1038/s42003-022-03505-7

Dentico, D., Cheung, B. L., Chang, J. Y., Guokas, J., Boly, M., Tononi, G., & Van Veen, B. (2014). Reversal of cortical information flow during visual imagery as

compared to visual perception. *NeuroImage*, *100*, 237–243. https://doi.org/10.1016/j.neuroimage.2014.05.081

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

Duan, K., Xie, S., Zhang, X., Xie, X., Cui, Y., Liu, R., & Xu, J. (2023). Exploring the Temporal Patterns of Dynamic Information Flow during Attention Network Test (ANT). *Brain Sciences*, *13*(2). https://doi.org/10.3390/brainsci13020247

ElShafei, H. A., Masson, R., Fakche, C., Fornoni, L., Moulin, A., Caclin, A., & Bidet-Caulet, A. (2022). Age-related differences in bottom-up and top-down attention: Insights from EEG and MEG. *European Journal of Neuroscience*, 55(5), 1215–1231. https://doi.org/10.1111/ejn.15617

Feng, E & Crooks, G. (2008). Length of Time's Arrow. *Physical review letters*. 101. 090602. https://doi.org/10.1103/PhysRevLett.101.090602

Friston K. (2010). The free-energy principle: a unified brain theory?. *Nature reviews. Neuroscience*, *11*(2), 127–138. https://doi.org/10.1038/nrn2787

Gazzaley, A., Cooney, J. W., McEvoy, K., Knight, R. T., & D'Esposito, M. (2005). Top-down enhancement and suppression of the magnitude and speed of neural activity. *Journal of cognitive neuroscience*, *17*(3), 507–517. https://doi.org/10.1162/0898929053279522

G-Guzmán Elvira, Perl Yonatan Sanz, Vohryzek Jakub, Escrichs Anira, Manasova Dragana, Türker Başak, Tagliazucchi Enzo, Kringelbach Morten, Sitt Jacobo D. and Deco Gustavo. (2023). The lack of temporal brain dynamics asymmetry as a signature of impaired consciousness states. *Interface Focus*. 13:20220086. https://doi.org/10.1098/rsfs.2022.0086

Goddard, E., Carlson, T. A., & Woolgar, A. (2022). Spatial and Feature-selective Attention Have Distinct, Interacting Effects on Population-level Tuning. *Journal of cognitive neuroscience*, *34*(2), 290–312. https://doi.org/10.1162/jocn_a_01796

Gnesotto, F. S., Mura, F., Gladrow, J., & Broedersz, C. P. (2018). Reports on Progress in Physics Broken detailed balance and non-equilibrium dynamics in living systems: a review Broken detailed balance and non-equilibrium dynamics in living systems: a review Review. *Rep. Prog. Phys*, *81*, 32. Article 066601. https://doi.org/10.1088/1361-6633/aab3ed

Green, D. M., & Swets, J. A. (1966). Signal detection theory and psychophysics. John Wiley.

Hanslmayr, S., Aslan, A., Staudigl, T., Klimesch, W., Herrmann, C. S., & Bäuml, K. H. (2007). Prestimulus oscillations predict visual perception performance between and within subjects. *NeuroImage*, *37*(4), 1465–1473. https://doi.org/10.1016/j.neuroimage.2007.07.011

Harris, K. D., & Thiele, A. (2011). Cortical state and attention. *Nature reviews. Neuroscience*, *12*(9), 509–523. https://doi.org/10.1038/nrn3084

Iemi, L., Chaumon, M., Crouzet, S. M., & Busch, N. A. (2017). Spontaneous Neural Oscillations Bias Perception by Modulating Baseline Excitability. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, *37*(4), 807–819. https://doi.org/10.1523/JNEUROSCI.1432-16.2016

Jas, M., Engemann, D. A., Bekhti, Y., Raimondo, F., & Gramfort, A. (2017). Autoreject: Automated artifact rejection for MEG and EEG data. *NeuroImage*, *159*, 417–429. https://doi.org/10.1016/j.neuroimage.2017.06.030

Jelic, V., & Darreh-Shori, T. (2010). Donepezil: A Review of Pharmacological Characteristics and Role in the Management of Alzheimer Disease. *Clinical*

Medicine Insights: Therapeutics, *2*, CMT.S5410. https://doi.org/10.4137/CMT.S5410

Jensen, O., & Mazaheri, A. (2010). Shaping functional architecture by oscillatory alpha activity: gating by inhibition. *Frontiers in human neuroscience*, *4*, 186. https://doi.org/10.3389/fnhum.2010.00186

Kringelbach, M.L., Sanz Perl, Y., Tagliazucchi, E., Deco, G., (2023). Toward naturalistic neuroscience: Mechanisms underlying the flattening of brain hierarchy in movie-watching compared to rest and task. *Sci. Adv.* 9, eade6049. https://doi.org/10.1126/sciadv.ade6049

Kringelbach, M. L., Sanz Perl, Y., & Deco, G. (2024). The Thermodynamics of Mind. *Trends in cognitive sciences*, 28(6), 568–581. https://doi.org/10.1016/j.tics.2024.03.009

Linkenkaer-Hansen, K., Nikulin, V. V., Palva, S., Ilmoniemi, R. J., & Palva, J. M. (2004). Prestimulus Oscillations Enhance Psychophysical Performance in Humans. *Journal of Neuroscience*, *24*(45), 10186–10190. https://doi.org/10.1523/JNEUROSCI.2584-04.2004

McCormick, D. A., Nestvogel, D. B., & He, B. J. (2020). Neuromodulation of Brain State and Behavior. *Annual Review of Neuroscience*, *43*(Volume 43, 2020), 391–415. https://doi.org/10.1146/annurev-neuro-100219-105424

McGinley, M. J., David, S. V., & McCormick, D. A. (2015). Cortical Membrane Potential Signature of Optimal States for Sensory Signal Detection. *Neuron*, *87*(1), 179–192. https://doi.org/10.1016/j.neuron.2015.05.038

Montefusco-Siegmund, R., Schwalm, M., Rosales Jubal, E., Devia, C., Egaña, J. I., & Maldonado, P. E. (2022). Alpha EEG Activity and Pupil Diameter Coupling during Inactive Wakefulness in Humans. *eNeuro*, *9*(2), ENEURO.0060-21.2022. https://doi.org/10.1523/ENEURO.0060-21.2022

Nuiten, S. A., De Gee, J. W., Zantvoord, J. B., Fahrenfort, J. J., & van Gaal, S. (2023). Catecholaminergic neuromodulation and selective attention jointly shape perceptual decision-making. *Elife*, 12, RP87022. https://doi.org/10.7554/eLife.87022.2

Pfeffer, T., Ponce-Alvarez, A., Tsetsos, K., Meindertsma, T., Gahnström, C. J., van den Brink, R. L., … Donner, T. H. (2021). Circuit mechanisms for the chemical modulation of cortex-wide network interactions and behavioral variability. *Science Advances*, *7*(29), eabf5620. https://doi.org/10.1126/sciadv.abf5620

Pilipenko, A., & Samaha, J. (2024). Double Dissociation of Spontaneous Alpha-Band Activity and Pupil-Linked Arousal on Additive and Multiplicative Perceptual Gain. *Journal of Neuroscience*, *44*(19). https://doi.org/10.7554/eLife.68265

Podvalny, E., King, L. E., & He, B. J. (2021). Spectral signature and behavioral consequence of spontaneous shifts of pupil-linked arousal in human. *eLife*, *10*, e68265. https://doi.org/10.7554/eLife.68265

Safron, A., Klimaj, V, & Hipólito, I. (2022). On the importance of being flexible: Dynamic brain networks and their potential functional significances. *Frontiers in Systems Neuroscience.* 15 (2022), 149. https://doi.org/10.3389/fnsys.2021.688424

Samaha, J., Iemi, L., Haegens, S., & Busch, N. A. (2020). Spontaneous Brain Oscillations and Perceptual Decision-Making. *Trends in cognitive sciences*, *24*(8), 639–653. https://doi.org/10.1016/j.tics.2020.05.004

Sarter, M., & Lustig, C. (2020). Forebrain Cholinergic Signaling: Wired and Phasic, Not Tonic, and Causing Behavior. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, *40*(4), 712–719. https://doi.org/10.1523/JNEUROSCI.1305-19.2019

Shinozuka, K., Tewarie, P. K. B., Luppi, A., Lynn, C., Roseman, L., Muthukumaraswamy, S., Nutt, D.J., Carhart-Harris, R., & Kringelbach, M. L.

(2024). LSD flattens the hierarchy of directed information flow in fast whole-brain dynamics. *bioRxiv*. https://doi.org/10.1101/2024.04.25.591130

Smulders, F.T.Y., ten Oever, S., Donkers, F.C.L., Quaedflieg, C.W.E.M. and van de Ven, V. (2018), Single-trial log transformation is optimal in frequency analysis of resting EEG alpha. Eur J Neurosci, 48: 2585-2598. https://doi.org/10.1111/ejn.13854

Twomey, D.M., Murphy, P.R., Kelly, S.P. and O'Connell, R.G. (2015), The classic P300 encodes a build-to-threshold decision variable. Eur J Neurosci, 42: 1636-1643. https://doi.org/10.1111/ejn.12936

Ulanowicz, R. & Hannon, B. (1987). *Life and the production of entropy*. Proc Royal Soc London. Proceedings of The Royal Society of London. Series B, *Biological Sciences* (1934-1990). 232. 181-192. https://doi.org/10.1098/rspb.1987.0067

van Vugt, M. K., Beulen, M. A., & Taatgen, N. A. (2019). Relation between centro-parietal positivity and diffusion model parameters in both perceptual and memory-based decision making. *Brain research*, *1715*, 1-12. https://doi.org/10.1016/j.brainres.2019.03.008

Waschke, L., Tune, S., & Obleser, J. (2019). Local cortical desynchronization and pupil-linked arousal differentially shape brain states for optimal sensory performance. *eLife*, *8*, e51501. https://doi.org/10.7554/eLife.51501

Welch, P. (1967). The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms. *IEEE Trans. Audio Electroacoustics* 15, 70–73. https://doi.org/10.1109/TAU.1967.1161901

Yerkes RM, Dodson JD (1908). "The relation of strength of stimulus to rapidity of habit-formation". *Journal of Comparative Neurology and Psychology*. **18** (5): 459–482. https://doi.org/10.1002%2Fcne.920180503

Statement of Contribution

Conceptualization: Jake Tear, Elvira Del Agua, Stijn Nuiten, Simon Van Gaal, Gustavo Deco

Methodology Development: Gustavo Deco, Stijn Nuiten, Simon Van Gaal

Software Programming: Jake Tear, Elvira Del Agua, Stijn Nuiten

Research Procedure: Stijn Nuiten, Simon Van Gaal

Formal analysis: Jake Tear, Elvira Del Agua

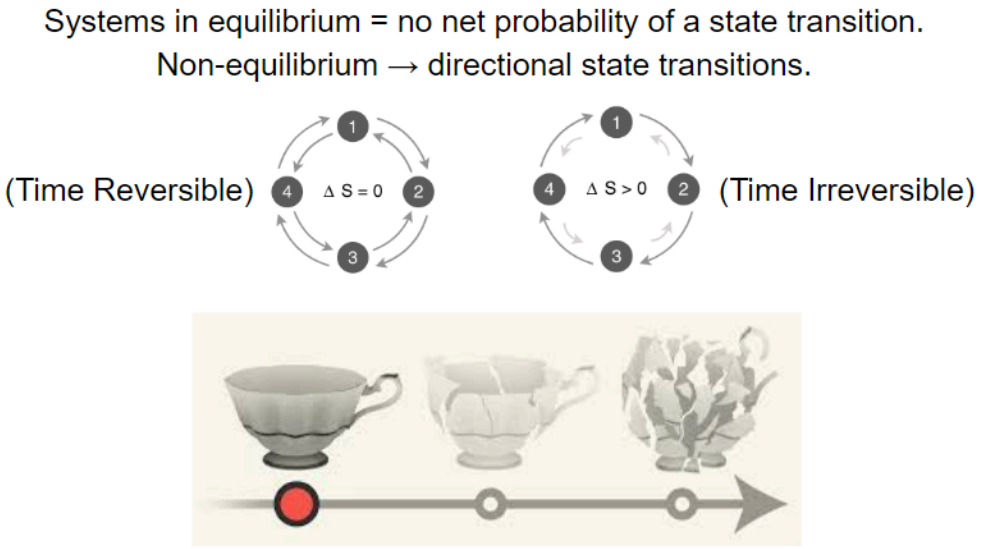
Visualization: Jake Tear

Preparation of the original draft: Jake Tear

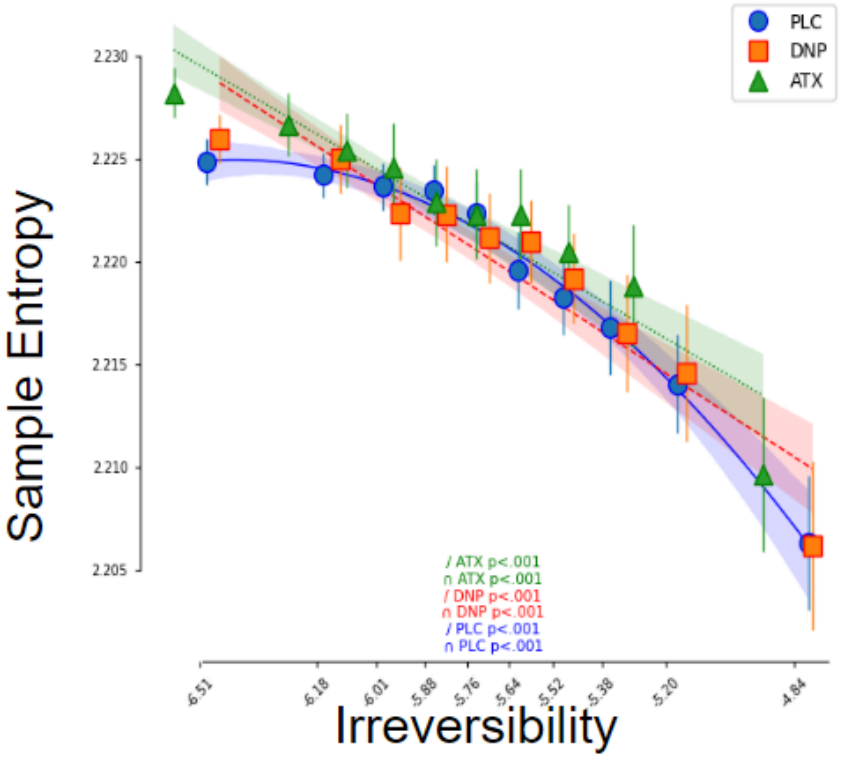
Editing: Jake Tear, Elvira Del Agua, Gustavo Deco

Supplemental Figures

Supplemental Figure 1

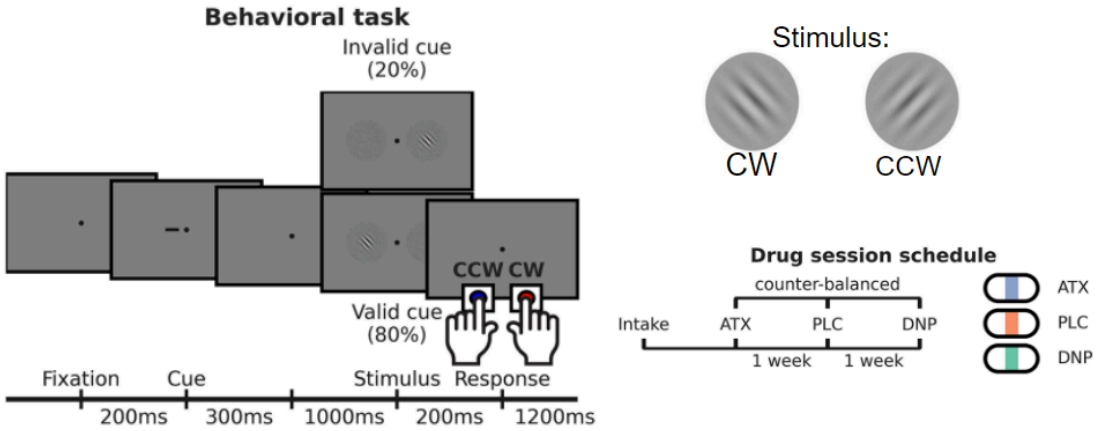


Supplemental Figure 2

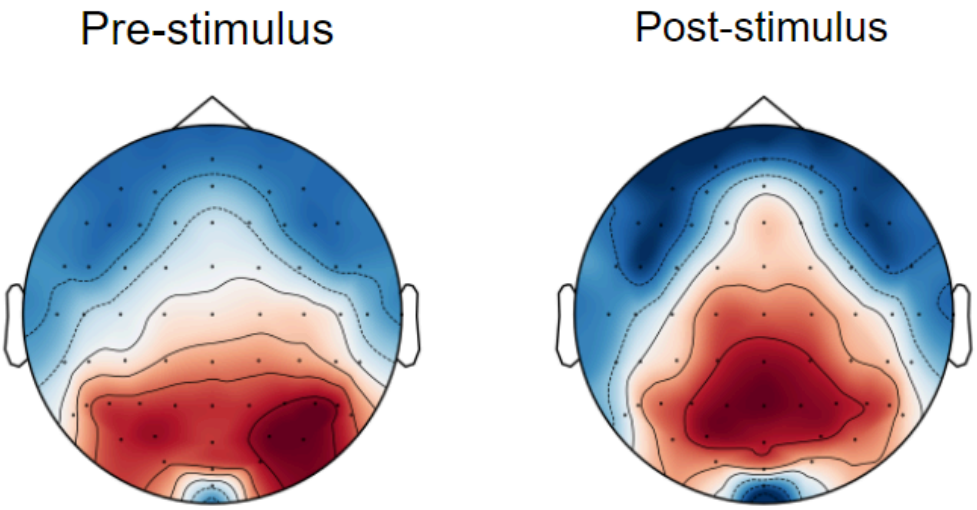


Theoretically, in equilibrium, entropy is maximal and entropy production is minimal. Here we see this empirically as well.

Supplemental Figure 3



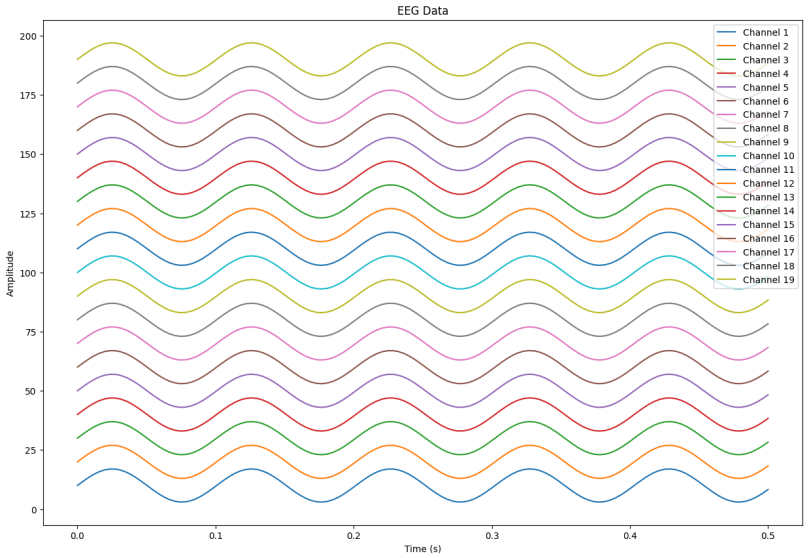
Supplemental Figure 4



Regional Irreversibility values of pre- and post-stimulus plotted in topographic form.

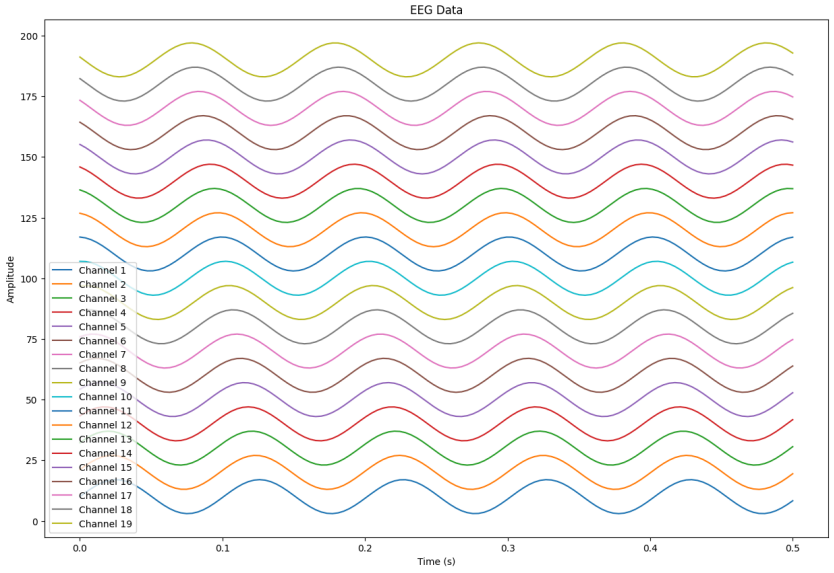
Supplemental Figure 5

When all oscillatory phases align:



The signal is completely reversible.

When phases are linearly incremented:



The signal is highly irreversible.