Integrated information phi is reduced in anaesthetised flies

Angus Leung

Word Count:

A paper submitted in partial fulfilment of the requirements of the degree of

*Bachelor of Science (Honours)*

*School of Psychological Sciences, Monash University*

*[Month, 2017]*Table of Contents

[Abstract 4](#_Toc488651264)

[Statement of Contribution 5](#_Toc488651265)

[CHAPTER 1: INTRODUCTION 6](#_Toc488651266)

[Overview (is a title needed?) 6](#_Toc488651267)

[Integrated Information Theory 7](#_Toc488651268)

[Integrated Information from the Decoding Perspective (include as subheading under IIT?) 9](#_Toc488651269)

[Mutual Information 9](#_Toc488651270)

[Conditional Mutual Information 9](#_Toc488651271)

[Loss of Consciousness and Feedback Disruption Under Anaesthesia 9](#_Toc488651272)

[LFP 9](#_Toc488651273)

[Aims and Hypotheses (in a separate section?) 9](#_Toc488651274)

[CHAPTER 2: EXPERIMENTAL METHODS AND RESULTS 11](#_Toc488651275)

[Method 11](#_Toc488651276)

[Experimental Procedure 11](#_Toc488651277)

[Animal preparation. 11](#_Toc488651278)

[Electrode probe insertion. 11](#_Toc488651279)

[Isoflurane delivery. 12](#_Toc488651280)

[Experimental protocol. 12](#_Toc488651281)

[Local field potential preprocessing. 12](#_Toc488651282)

[Φ3 Computation 13](#_Toc488651283)

[Discretisation. 13](#_Toc488651284)

[Network Selection. 13](#_Toc488651285)

[Transition Probability Matrix Construction. 13](#_Toc488651286)

[Φ calculation. 14](#_Toc488651287)

[Φ\* Computation 14](#_Toc488651288)

[Data Analysis 15](#_Toc488651289)

[Results 17](#_Toc488651290)

[Phi-3 greater during wakefulness 17](#_Toc488651291)

[Equal likelihood of MIP cuts 18](#_Toc488651292)

[Moderate correlation with phi\* 19](#_Toc488651293)

[CHAPTER 3: DISCUSSION AND CONCLUSION 21](#_Toc488651294)

[CHAPTER 4: REFERENCES 22](#_Toc488651295)

# Abstract

Blardiblarblar

# Statement of Contribution

|  |  |
| --- | --- |
| Signed: |  |
|  | Angus Leung |

Date: /2017

# **CHAPTER 1: INTRODUCTION**

30% of the word count is for background/literature review

## Overview (is a title needed?)

Opening statement - start with concept of consciousness – hard problem of consciousness, mention that there are existing theories of consciousness, one which somewhat tackles the hard problem is IIT. The aim of this project is threefold: to test if IIT’s proposed measure is larger during consciousness (or conversely reduced under anaesthesia), to compare it’s latest, computationally expensive derivation to a more practical version, and to compare the derivation of phi to a past finding in the same data. This project aims to investigate the validity of IIT’s measure of consciousness by applying it to local field potentials recorded from the fly brain.

A key focus in neuroscientific research on consciousness has been to find how consciousness arises from neural activity in the brain. The prevalent method with which to tackle this question has been to search for neural activities which correlate with consciousness – the search for the neural correlates of consciousness (NCC; {Koch, 2016 #12}). Following this approach, many specific neural interactions have been proposed as potential NCC, such as synchronous activation among neurons, or feedback interactions. Whether such correlations are reliable indicators is however debatable. For example, synchronous activity and feedback both occur in the cerebellum, which likely does not directly contribute to consciousness. Furthermore, they lose relevance as indicators of consciousness in contexts involving brain damage, non-human animals, and artificial systems, and as correlates they fail to provide an explanation as to how consciousness arises. In order to understand how consciousness arises, testable theories which address what consciousness is and what physical interactions it requires are needed. This research investigates the validity of one such theory: the integrated information theory of consciousness.

## The Integrated Information Theory of Consciousness

A relatively recent approach to consciousness (compared to others) is the integrated information theory

Information is what we ‘get out’ of an event which occurs. If the probability of some event occurring is 1 (i.e. it always occurs), then the occurrence of that event gives no information (as we already know that it always occurs). In this case, the information we get from the event occurring is 0. This is the lower bound of information (however, some advocate for the utility of negative information)

Entropy is a measure of uncertainty. For example, a coin flip has two possible outcomes. With two possible outcomes, uncertainty (i.e. entropy) would be maximised if the coin was unbiased – we wouldn’t be able to predict the outcome any better than chance. If the coin was biased however, say towards heads, then the outcome is less uncertain as it is more likely to be heads. In the extreme case where the outcome is always heads, there would be no uncertainty at all – in this case there is only really one outcome. If we know that the coin is biased in such a manner, observing the outcome would give us no information (as we practically already know the outcome). If we knew that the coin was slightly biased, we would get a little information. Conversely, if the coin was fair, then observing the outcome would give us information, i.e. the outcome. It is intuitive that the more ‘uncertain’ a set of events is, the more information we get from any of the events occurring.

Conditional entropy of some variable is the entropy of that variable given a fixed value of another variable.

IIT posits that integrated information phi is consciousness. To understand what is meant by “integrated information”, perhaps it is a good idea to break it into two parts: information and integration. Information may be understood as a reduction in uncertainty. As an example, let’s consider a system of two simple neurons. At any one timepoint, each of the two simple neurons is either firing or not firing. Thus at any timepoint, the system takes one of four states, as illustrated in Figure. If we know or consider nothing more about the two neurons, at any timepoint t we could only guess their state at chance accuracy (i.e. in this scenario we would have a one in four chance of guessing correctly), even if we knew their past state (at t-1) or even future state (at t+1). This uncertainty is referred to as entropy. However, if we knew that second of the neurons at time t always assumed the state of the first neuron at time t-1, considering the past system state would essentially allow us to eliminate two possibilities for time t, increasing the chances of a correct guess (which would now be one in two). This new uncertainty, given a fixed condition (knowing the first neuron at t-1) is referred to as the conditional entropy (the entropy conditioned upon a fixed event). The reduction of uncertainty by considering the past state (consequently increasing our guess chance from ¼ to ½) is referred to as mutual information. In our specific case, because our information is derived from entropies before and after taking into account the past state, it is formally referred to as mutual information (or information gain – mutual information and information gain are equivalent).

The modifier “integrated” specifies that any information gain (from considering the system state at some other time) should be due to dependencies among elements within the system. A simple example to demonstrate this: consider two brains as a single system. While considering the past of both brains allows us to better predict the present states within both brains, we can also consider the two brains independently and predict just as better. In other words, considering the two brains together gives us no additional information than just considering one at a time, and so there is no integration and thus no integrated information.

The last iteration of IIT (3.0) updated both of how to calculate information and how to assess integration. Information is now assessed using the earth mover’s distance instead of Kullback-Leibler divervence (with KL-divergence being the generally accepted distance measure, but not metric, in information theory), and integration is now considered in a bidirectional manner.

A key limitation of the measure proposed by IIT is in computational complexity. As the number of elements increases, compute time grows exponentially. This makes it impractical to directly apply it to a system as complex as the human brain.

## Integrated Information from the Decoding Perspective (include as subheading under IIT?)

### Mutual Information

### Conditional Mutual Information

## Loss of Consciousness and Feedback Disruption Under Anaesthesia in Flies (move to between IIT 3.0 and IIT 2.0) Flies as a Potential Model

The use of flies in place of humans reduces the weight of the limitation. In contrast to human brains which consist of x neurons, and rat brains which consist of r neurons, the typical fly brain consists of ~y neurons. In conjunction with this, the fly exhibits many behaviours which are easily controlled through gene manipulation.

## LFP (maybe not a necessary section)

## Aims and Hypotheses (in a separate section?)

Though it has been investigated in simulation studies, the latest formulation of phi provided by IIT 3.0 has not yet been empirically tested in a biological system. The first aim is thus to investigate and compare phi in the awake and anaesthetised fly. IIT predicts that phi will be reduced under anaesthesia. Given the past finding of stronger feedback influences during wakefulness which is reduced under anaesthesia, a sub aim is to replicate this finding using IIT, specifically by comparing MIP cuts between conditions. It is hypothesised that unidirectional cuts from centre channels to peripheral channels will be more likely under isoflurane. Finally, given the heavy computational cost of calculating phi, we also compare phi to a cheaper alternative in the hopes that the two will correlate. Once again, it is hypothesised that phi will be reduced under anaesthesia. …Thus, the primary aims of this project were as follows:

1. To investigate if phi behaves in a manner consistent with IITs predictions. Specifically, it is expected that phi will be reduced under anaesthesia, when compared to no anaesthesia.
2. To replicate the finding of reduced feedback under anaesthesia using a component of IIT, specifically MIP cuts. It is expected that unidirectional cuts from the centre of the brain to the periphery will be more likely under anaesthesia.
3. To compare phi with less computationally expensive potential measures of consciousness, specifically phistar and its components. It is expected that phistar will be correlated with phi, but not mutual information or partitioned mutual information.

# **CHAPTER 2: EXPERIMENTAL METHODS AND RESULTS**

## Method

### Experimental Procedure

The data used in this project is a subset of the data collected and preprocessed previously in {Cohen, #2}, where the full experiment is described. Here I only detail methods relevant to the dataset used in the present project.

Animal preparation. Thirteen female laboratory-reared Drosophila melanogaster flies (Canton S wild type, 3-7 days post eclosion) were collected under cold anaesthesia and tethered to a tungsten rod. Flies were glued dorsally to the rod using dental cement (Synergy D6 FLOW A3.5/B3, Coltène Whaledent) which was cured with blue light. The flies’ wings were also glued to the rod in order to prevent wingbeats during recording, and dental cement was applied to the neck to stabilise the head. Tethered flies were positioned above a 45.5 mg air-supported Styrofoam ball, setup similarly to {Paulk 2013}, and thus were able to walk in place.

Electrode probe insertion. Linear silicon probes with 16 electrodes (Neuronexus Technologies) were inserted laterally into the fly’s eye, perpendicular to its curvature, with the electrode recording sites facing posteriorly. Insertion was performed with the aid of a micromanipulator (Märzhäuser). Probes had an electrode site separation of 25 µm (3mm-35-177) and measured 375 µm from base to tip. As a reference electrode, a sharped fine tungsten wire (0.01 inch diameter, A-M Systems) was placed in the thorax. Recordings were made using a Tucker-Davis Technologies multichannel data acquisition system with a 25 kHz sampling rate. To ensure consistent probe insertion depth, probes were inserted until all electrodes were recording neural activity. This was confirmed by presenting a flickering visual stimulus (with spectral peak at 460 nm and 30nm half-peak width; flickering at 1 and 13 Hz), and subsequently observing steady state visually evoked potentials (SSVEPs) at the most peripheral electrode. The probe was then retracted until the most peripheral electrode showed little to no neural activity. Probe insertion in this manner does not seem to affect fly locomotion {Paulk 2013}.

Isoflurane delivery. Isoflurane was delivered from an evaporator (Mediquip) onto the fly through a connected rubber hose. The isoflurane was delivered at a constant flow of 2 l/min and continuously vacuumed from the opposite side of the fly. Actual concentration near the fly body was either 0 vol% (air condition) or 0.6 vol% (isoflurane condition) as estimated following the gas chromatography procedure described by {Kottler 2013} for measuring isoflurane concentration. Flies in the air condition responded to air puffs by moving their legs and abdomen, but were rendered inert under the isoflurane condition {Cohen, #2}.

Experimental protocol. The complete experimental procedure is described in {Cohen, #2}. Here I briefly describe the procedure relevant to the data used in this project. An experiment consisted of two blocks: one for the air condition, followed by one for the isoflurane condition. Each block started with a series of air puffs, followed by 18 s of rest, 248 s of visual stimuli, another 18 s of rest, and finally a second series of air puffs. Isoflurane was administered immediately after completion of the first block (i.e. after the last air puff), and flies were left for 180 s to adjust to the new concentration before beginning the second block. The data used in this project corresponds to the 18 s period between the end of the first series of air puffs and the beginning of the visual stimuli.

Local field potential preprocessing. LFPs were recorded at 25 kHz and downsampled to 1000 Hz. Electrodes were bipolar rereferenced by subtracting neighbouring electrodes, resulting in 15 signals. Hereafter these signals will be referred to as “channels”. The 18 s of data for each condition was split into 2.25s segments, giving 8 “trials” of 2250 samples each. Finally, line noise at 50 Hz was removed using the *rmlinesmovingwinc.m* function of the Chronux toolbox {<http://chronux.org/>; Mitra and Bokil, 2007} with three tapers, a windows size of 0.7 s, and a step size of 0.35 s.

### Φ3 Computation

The calculation of phi3 at any timepoint requires 1) a candidate network, 2) the state of the network (e.g. in a network consisting of binary elements, which elements are ‘on’ and which are ‘off’), and 3) the probabilities of transitioning from each system state to all other system states. In the context of this project, a candidate network consists of a set of channels, and its state is given by the discretised measurements of its channels. The Python 3.6.0 module in MASSIVE was used.

Discretisation. Discretisation of recordings was required as IIT 3.0 has yet to be extended to continuous variables. To account for this, I discretised the recordings of each channel using its median value. The median value of a channel was taken across samples using all eight trials at a single condition (air or iso). Samples were then replaced with a 1 if greater than the median, and a 0 otherwise.

Network Selection. Given that the time to calculate phi grows exponentially with the number of elements in a candidate system, candidate networks were limited to consisting of up to four channels (i.e. networks of 2, 3, or 4 channels). Within this limitation, all channel combinations were selected giving a total of 1830 candidate networks (15 choose 2 + 15 choose 3 + 15 choose 4). All networks were defined as fully connected, i.e. each channel was bidirectionally connected to every other channel. The state of a network at a given timepoint is given by the states of its channels (e.g. for channels A=1, B=0, C=1, the network state is 100).

Transition Probability Matrix Construction. A transition probability is the probability of a state at time t transitioning into another state at time t+tau. A transition probability matrix thus holds the probabilities of all states at time t transitioning into all other states at time t+tau. For a given network, source state, end state, and time lag tau, a transition probability was defined as the number of times across all trials the source state, after the time lag tau, transitioned into the end state divided by the total number of transitions (i.e. transitions to any state). As IIT’s exclusion postulate advocates for calculation at the optimal spatiotemporal resolution, transitions were calculated at three lag levels: 4, 8, and 16 ms.

Φ calculation. Phis was calculated using the PyPhi package for Python 3.6. The details of computing phi are provided in {Oizumi, 2014 #2}, so I here will give a brief summary of the calculation process. Phis were calculated using PyPhi’s compute.big\_mip function, which essentially takes a network and TPM as input, and provides a phi value as well as the unidirectional bipartition corresponding to the MIP. As each sample provides a state, the overall phi value for a trial of 2250 samples is the average of phi values, weighted by the number of occurrences of each state within the trial.

FIGURE: electrode insertion, rereferenced data, discretised data, tpm

### Φ\* Computation

Phistar was calculated in MATLAB 2016a, using a toolbox which implemented phistar calculation in a previous project {<https://github.com/amhaun01/phipattern>, Haun}. The details of computing phistar are given in {Oizumi, 2016 #5}, so once again I will provide only a summary of the process.

Phistar can be summarised relatively simply. It is the difference between the mutual information of a whole system, when considering all connections between its parts, and the mutual information of a partitioned system, where we ignore some set of connections in the original system. As inputs it takes covariances matrices corresponding to the covariance between channels at time t with the same channels at time t+tau, at time t with themselves, and at time t+tau with themselves, as well as a partitioning scheme. As phistar is based on version 2.0, it has the limitation of partitions with more independent groups having greater phi, thus to compare partitions a normalisation factors was applied. The partition returning the smallest normalised phi was taken as the MIP, and its unnormalized phi as the phistar value.

### Data Analysis

Phi-3. A linear mixed effects model was employed as an omnibus test for effects of tau lag, network size, and condition. Thus, the model included fixed effects of lag (4, 8, or 16 ms), network size (2, 3, or 4), and condition (air or iso). The nested random effect of networks being nested within flies was included by including random intercepts for fly and the interaction between fly and network. Due to heavy positive skew of phi values, trial averaged phi values were log transformed to address heteroscedasticity before fitting the model. Fixed effects were tested using simulated likelihood ratio tests (N = 1000) between the full model and a null model with the effect removed.

Feedback. To remain consistent with previous analysis on this data {Cohen, #2}, feedback was defined as an influence from a central channel to a periphery channel. Thus the following analysis was limited to networks consisting of two channels (will expand to 3 and 4 channels if there is room and time, following the definition that feedback = more feedback cuts than feedforward cuts in the MIP; may stick to using the same limited channels or use all channels). Following the scheme used in {Cohen, #2}, channels were grouped as either peripheral (channels 2-7) or central (channels 10-15). MIP cuts from a centre channel to a peripheral channel were considered as feedback cuts (and feedforward otherwise). As each sample gives a state and corresponding MIP, we took the portion of samples with a MIP with a feedback cut within a trial. A paired t-test was used to compare the trial averaged portions between conditions. (Check distribution, maybe a non-parametric test is valid here – distribution consists of values which are multiples of one-quarter).

Phi-\*. As with phi-3, a linear mixed effects model was used to assess the fixed effects of lag, network size, and condition in order to account for nesting. Once again, trial averaged values were log transformed to address heteroscedasticity. Simulated likelihood ratio tests (N=1000) comparing the full model with null models were used to test for fixed effects. Correlations between phi-star and phi-3 were calculated at each candidate network after averaged across trials. To assess MIP equivalence, directionality of IIT3 MIP cuts were ignored. Additionally, as IIT3 MIP cuts only bipartition the system, while IIT2 allows for partitions consisting of more than 2 sub-groups, trials in which the calculation of phi-star resulted in a MIP which was not a bipartition were not included in the analysis. MIPs were considered equal if each subgroup in the partition consisted of the same channels. As each trial results in only one phi-star MIP and multiple phi-3 MIPs (each sample takes one of n possible system states, with each state having its own MIP – note there is nothing to suggest that two states cannot have the same MIP). Thus each trial gives a portion of equal MIPs. T-tests were used to compare trial averaged portions between conditions (not corrected, as results are not significant at less conservative .05)

## Results

While it is unclear whether flies are “conscious” in that they have phenomenological experience, isoflurane reduces behavioural stimuli to noxious stimuli (perhaps this should be covered in the introduction, and if so, doesn’t need to be mentioned here? Or maybe just a reminder that responsiveness diminished under isoflurane?) Analyses were conducted using MATLAB 2017a and MATLAB 2015b (simulated likelihood ratio tests, via MASSIVE).

### Integrated information is reduced under isoflurane

Possible flow: find effect of condition, tau, and number of channels using omnibus LME, then to find pattern of differences use t-tests with correction

Possible flow: find pattern of differences using t-tests with correction, then confirm effect with omnibus LME (this doesn’t seem to work, as the omnibus test is generally conducted first, e.g. ANOVA following by post-hoc tests)

Descriptive statistics! Non-normal descriptives?

Due to the crossed nature of the data (channel combinations across flies), a linear mixed effects model with random intercepts for fly and the interaction between fly and channel combination was employed as an omnibus test for effects of tau lag, network size, and condition. Thus, the model included fixed effects of lag (4, 8, or 16 ms), network size (2, 3, or 4), and condition (air or iso) and random intercepts for fly and the interaction between fly and channel combination. To address heteroscedasticity due to heavy positive skew of phi values, trial averaged phi was log transformed before fitting the model. Fixed effects were tested using simulated likelihood ratio tests (N = 1000) between the full model and a null model without the effect. There was a significant effect of lag (stats), with longer lags giving smaller phi values, as well as of network size (stats), with larger networks giving larger phis. Importantly, condition was a significant effect (stats), with the isoflurane condition giving reduced phis in comparison to the air condition. (maybe include table of LR stats and coeffs, with note that coeffs are for log transformed phi – is it necessary to report on coefficients, and if so, report coeffs for transformed data, inverse transformed coeff after fitting to transformed data, or coeffs for untransformed data?). Figure X shows the averaged phi values and delta (air – iso) phi values for every candidate network (at params). Post-hoc paired t-tests (with FDR correct p < .05) suggest that significant differences are more likely for networks consisting of more central channels. This is more evident for longer tau. Maybe include some proportions, e.g. sig/nonsig at each network size

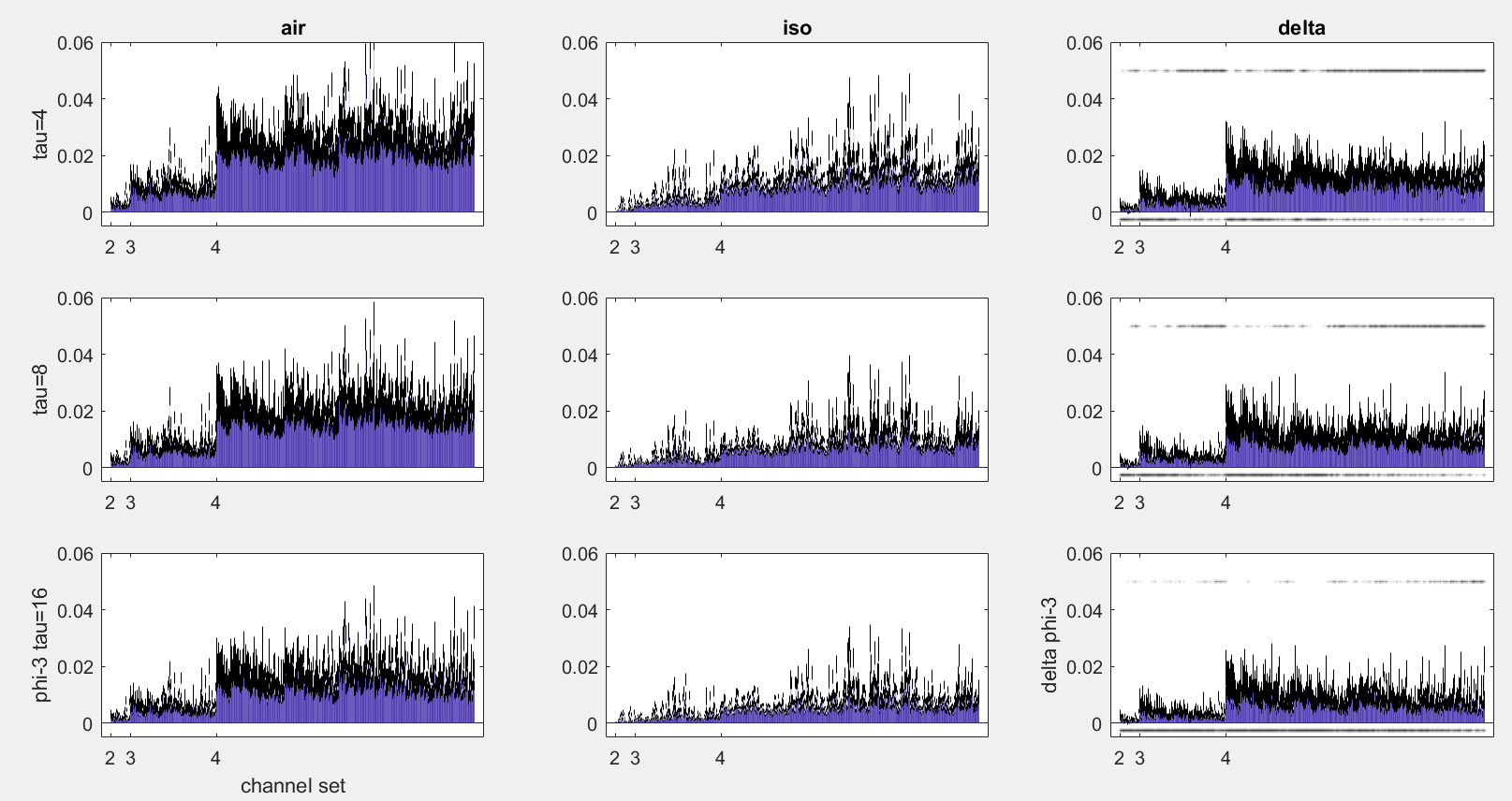


FIGURE: averaged phis for all channel sets: two columns for two taus (extremes: 4 and 16 ms), rows are air, iso, and delta; significance pattern must be visible (to convey possible point that more central combinations give a significant difference)

### Equal likelihood of feedback MIP cuts under isoflurane

As it is not immediately obvious as to whether a partitioning of three channels or more is feedback or not, we took only candidate networks consisting of 2 channels and compared the portion of feedback to feedforward cuts between conditions. This parallels the two-channel nature of GC analysis which was previously conduced on the data {Cohen, #2}. For comparability with the past finding of reduced feedback in the data under iso, the same periphery-centre channel pairings were selected.

FIGURE: boring bar plot – maybe conduct tests per channel set, like in other sections

### Phi-star is moderately correlated with phi-3

As with phi-3, an linear mixed effects model was used to assess the fixed effects of condition, lag, and channels used. Figure x shows the values at each channel combination (no t-tests survived correction for multiple comparisons at q=0.05; logged t-tests still give results). This was repeated for mutual information and entropy.

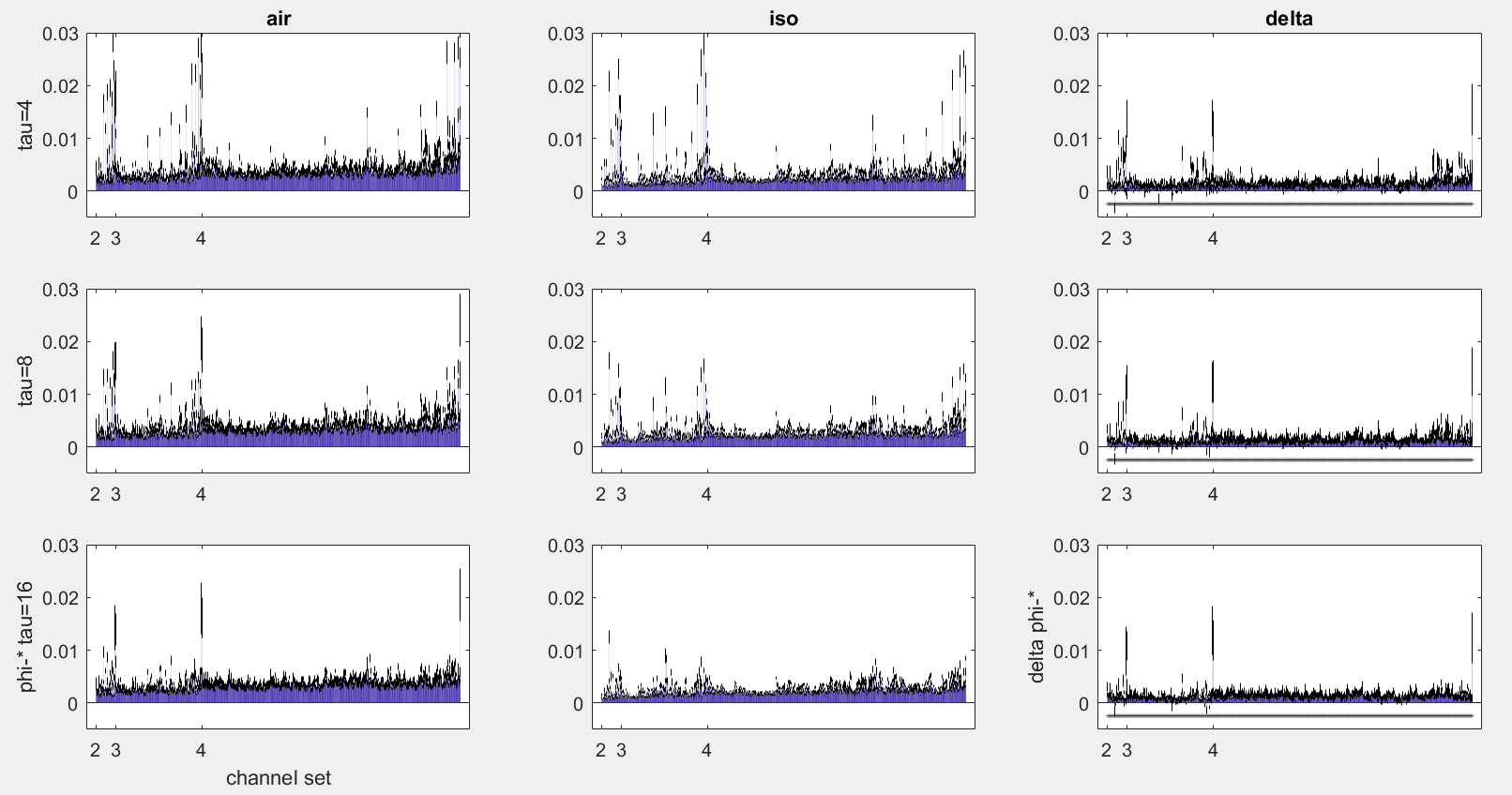


FIGURE: phistars at two taus, same as phi3 figure

Despite no difference between conditions at the individual network level, phi-star at individual networks was generally correlated with phi-3 (after correction). More correlations were significant in the air condition, and the proportion of significant correlations increased with tau lag. (Look into proportion of significance per nChannels). The average correlation (after Fisher z-r transformation and backtransform) was x.

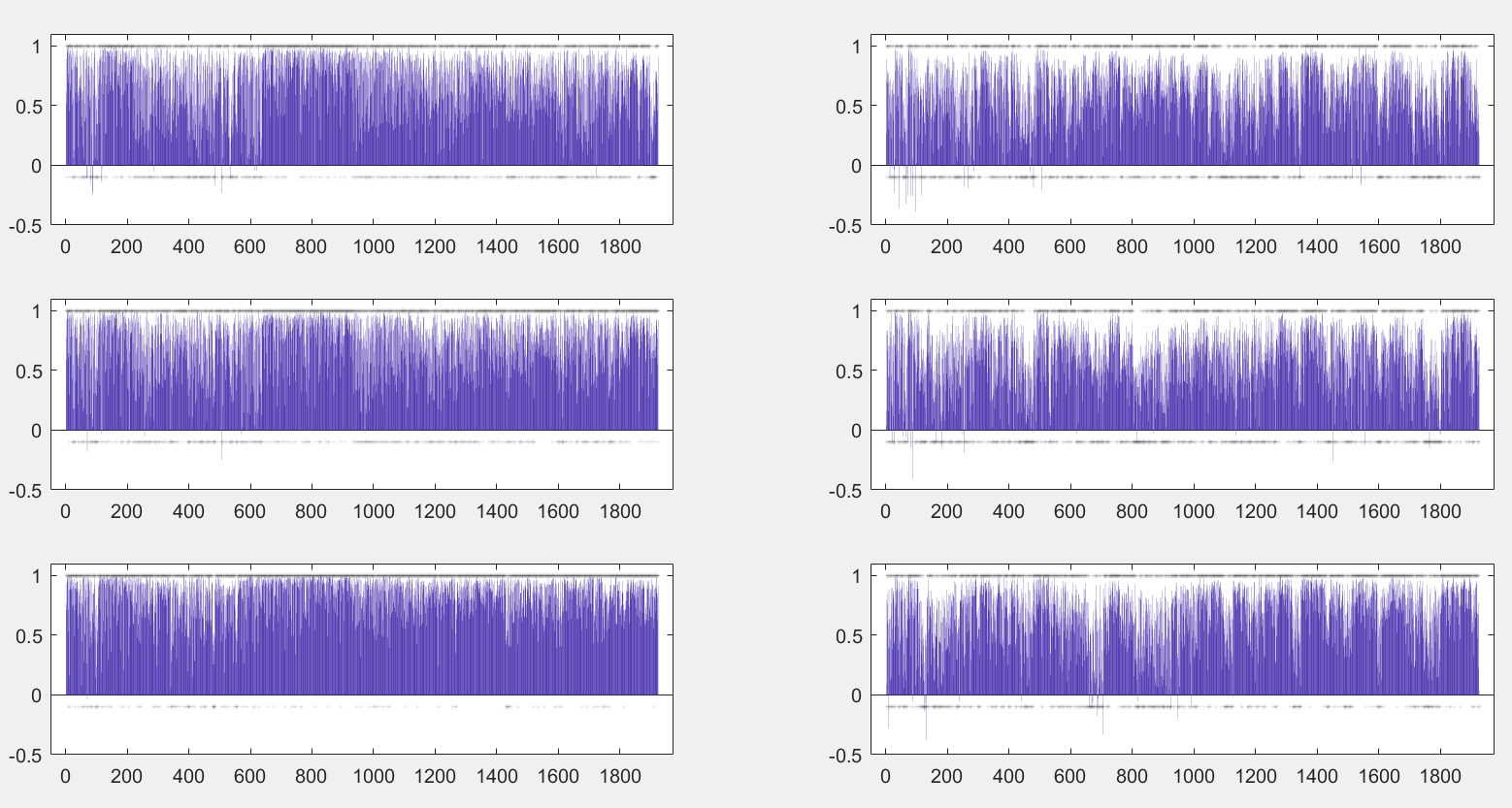


FIGURE: phi correlations at one tau, similar as phi3 figure

However, the likelihood of having matching MIPs was not significantly different from chance.

## CHAPTER 3: DISCUSSION AND CONCLUSION

Blah blah blah discuss discuss, etc. etc.

More blah blah blah stuff.

This stuff can be related to modelling the brain as a graph, possible future approach may be to calculate phi in the core vs in the periphery. The two approaches (IIT and core-periphery search) may go together as IIT calculates phi over a set of nodes. Furthermore the search for phi may help identify the core (or at least the conscious core), which may be dynamic, especially in line with the ideas of segregation and integration.

A major direction is in the algorithmic/mathematical derivation of phi. A key limiting factor to computing phi is the search for the MIP, which requires searching over all possible partitions of a system. Reduction of this problem to decrease compute time is already underway, but given the mathematical nature of the theory proofs are required equating approximations of the MIP to the actual MIP.

## CHAPTER 4: REFERENCES