Integrated information phi in flies reduces under anaesthesia

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 2](#_Toc490574567)

[Statement of Contribution 2](#_Toc490574568)

[CHAPTER 1: INTRODUCTION 2](#_Toc490574569)

[The Integrated Information Theory of Consciousness 2](#_Toc490574570)

[The Fly Model for Studying Consciousness **Error! Bookmark not defined.**](#_Toc490574571)

[Aims and Hypotheses (in a separate section?) 2](#_Toc490574572)

[CHAPTER 2: EXPERIMENTAL METHODS AND RESULTS 2](#_Toc490574573)

[Method 2](#_Toc490574574)

[Experimental Procedure 2](#_Toc490574575)

[Animal preparation. 2](#_Toc490574576)

[Electrode probe insertion. 2](#_Toc490574577)

[Isoflurane delivery. 2](#_Toc490574578)

[Experimental protocol. 2](#_Toc490574579)

[Local field potential preprocessing. 2](#_Toc490574580)

[Φ Computation 2](#_Toc490574581)

[Discretisation. 2](#_Toc490574582)

[Network Selection. 2](#_Toc490574583)

[Transition Probability Matrix Construction. 2](#_Toc490574584)

[Φ calculation. 2](#_Toc490574585)

[Φ\* Computation 2](#_Toc490574586)

[Data Analysis 2](#_Toc490574587)

[Phi-3. 2](#_Toc490574588)

[Feedback. **Error! Bookmark not defined.**](#_Toc490574589)

[Phi-\*. 2](#_Toc490574590)

[Results 2](#_Toc490574591)

[Integrated information is reduced under isoflurane 2](#_Toc490574592)

[Equal likelihood of feedback MIP cuts under isoflurane **Error! Bookmark not defined.**](#_Toc490574593)

[Phi-star is moderately correlated with phi-3 2](#_Toc490574594)

[CHAPTER 3: DISCUSSION AND CONCLUSION 2](#_Toc490574595)

[CHAPTER 4: REFERENCES 2](#_Toc490574596)

# Abstract

The integrated information theory of consciousness proposes a potential measure of conscious level, integrated information Φ, which is predicted by the theory to be high during consciousness and low during loss of conscious level. The latest derivation of the quantity however has issues regarding practical applicability to large systems, and has only been utilised in simulation studies. Consequently, the timescale for which Φ is maximal is also unclear. Thus, this project aimed to (a) assess the construct validity of Φ in a real biological system, (b) assess Φ at varying timescales, and (c) compare it to a potentially more practical derivation of integrated information, Φ\*. To achieve these aims, I calculated both Φ and Φ\* across sets of recordings obtained from the fly brain during and without administration of isoflurane anaesthesia. Both Φ and Φ\* were significantly reduced during anaesthesia.

# Statement of Contribution

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

Date: /2017

# **CHAPTER 1: INTRODUCTION**

A key focus in neuroscientific research on consciousness has been to find how consciousness arises from neural activity in the brain. In this context, consciousness refers to subjective experience, or “what it is like to be” {Thomas Nagel}. The primary approach with which to tackle this question has generally been to search for neural activities and interactions which correlate with consciousness – to search for the neural correlates of consciousness (NCC; {Koch, 2016 #12}). However, correlates alone cannot provide an explanation as to how consciousness arises. To understand how consciousness comes to be, testable theories which address what consciousness is and what physical interactions it requires are needed. Integrated information theory (IIT; {Oizumi, 2014 #45}) is one theory which has risen to provide a such a principled account of consciousness.

## The Integrated Information Theory of Consciousness

<Philosophical stuff>Consciousness is one hard problem. Conceptually, it is the capacity to have subjective experience, or, to paraphrase Thomas Nagel, it is “what it is like to be”. This idea of consciousness is distinct from other terms {from Klein 2016} which might also be termed consciousness, such as self-awareness/self-consciousness {}, high-order thoughts {}, or reportable access to one’s own experience {}. Thus, a system which is conscious has an experience “of itself”, and a system which is unconscious does not. Despite being just as prevalent as any other phenomenon (arguably more prevalent, as the only way to be experience any phenomenon is to, well, experience it), we still lack an understanding of how consciousness arises from physical neural activity.

<Prevalence of reductionism>The prevalent method with which to approach this question has been to search for and study the NCC. Within this approach, researchers focus either only on finding NCCs for levels of consciousness (the minimally sufficient conditions for a creature’s overall conscious level, such as wakefulness as opposed to dreamless sleep) or for contents of consciousness (the minimally sufficient conditions for a specific conscious perception, such as of that of a face or the colour red). This search has led to the proposal of numerous specific neural interactions as potential NCC, such as synchronous activation among neurons {Engel, 2001 #18}, 40 Hz oscillations in the cerebral cortex {Llinás, 1993 #58}, or feedback interactions {Lamme, 2010 #30} for the level of consciousness, and activation of the fusiform face area during face perception {Pierce, 2001 #57} for the contents of consciousness. However, the reliability of these correlates as indicators of consciousness is debatable. For example, synchronous activity and feedback interactions both occur in the cerebellum {Person, 2012 #19;Witter, 2016 #29}, which likely does not contribute to consciousness {Yu, 2015 #20}, while the fusiform face area is activated during perception of non-face stimuli {see PSY4270 essay}. Additionally, these correlates lose relevance in contexts involving brain damage, non-human animals, and artificial systems. It is also increasingly clear that this split approach to studying consciousness has significant drawbacks. Namely, the search for content NCCs presumes consciousness in a system, and the search for level NCCs … {Hohwy, 2009 #32}. More critically, while informative as to the kind of structures and interactions which cause consciousness, correlates leave an explanatory gap between physical interactions and the rise of consciousness. To address this gap, a theory of consciousness which identifies the key principles of consciousness and how they are achieved through interactions is required.

<Discussion of other theories>… Though the search for correlates in this manner continues, given the limitations of NCCs there has recently been a stronger emphasis on approaching the problem using a more theoretical perspective in order to target both level and contents simultaneously {Hohwy, 2009 #32}. Accordingly, a number of theories of consciousness have been proposed, such as the global workspace theory {Baars, 1997 #4;Baars, 2002 #5}, cross-order integration theory {Kriegel, 2007 #59}, … A common concept in such theories is the concept of integration across parts. For example, the global workspace theory advocates that widespread activation across distant brain regions is indicative of “integration” of “information”. Though they provide predictions as to what kinds of biological systems and interactions might give rise to consciousness, they however leave terms such as integration and information abstract and unoperationalised, resulting in an inability to provide any measure of the consciousness they hope to explain. Another weakness of such theories is their approach of building explanations for the generation of consciousness from interactions which have been identified as potential NCCs. Consequently, the explanations they propose are also limited to vertebrates.

<Intro to IIT>IIT stands out from the other theories by taking a different approach. Instead of building a theory from observed neural activities, IIT identifies fundamental aspects of consciousness, and from these reasons the necessary mechanisms for it. The fundamental properties it identifies are as follows: (a) intrinsic existence: an experience exists, and furthermore it exists from its own intrinsic perspective, independent of external observers; (b) composition: an experience is composed of multiple aspects (for example, the experience of watching a movie is composed of vision and audition, and the experience of a face is composed of eyes, a nose, etc.); (c) information: an experience rules out every other possible experience that could be occurring; (d) integration: an experience exists as a single whole which cannot be broken into parts; and (e) exclusion: an experience cannot be superposed with other experiences (consciousnesses cannot overlap). From these principles, it builds a measure, integrated information Φ, whose magnitude and structure are equivalent respectively to the level of consciousness in a system and its contents.

<Concept of phi>The raw concept of Φ can be understood from the principles of information and integration. Information (I) can be understood as a reduction in uncertainty. For example, a simple neuron at any time may have one of two states: firing or not firing. As the neuron can take one of multiple (two) states, there is inherent uncertainty as to which state it is in at any time. However, if we were to consider a second simple neuron and its connection with the first (e.g. the first neuron fires at time *t* if the second neuron was firing at *t* - 1), knowing the state of the second neuron at time *t* - 1 would allow us to deduce the state of the first neuron at *t*, reducing the set of possible states it could take at that timepoint (from two states to one). Similarly, knowing the state of the first neuron at time *t* would reduce the possible states of the second neuron at time *t* - 1. This reduction in uncertainty is information. Integration (O) refers to what exists only when considering a system as a whole - in this example, the two neurons and their connection. If the system of two neurons is split into independent parts, each consisting of a single neuron, then the uncertainty as to the state of the first neuron at any time *t* increases to what it was before we considered the second neuron (and vice versa for the second neuron). Thus, the information gained from considering the second neuron only exists when considering the whole system, and is considered to be the result of integration.

<Intrinsic, past+future, large system level>In the novel example provided, information regarding one part of the system (neuron one) is gained when given knowledge about another part (neuron two). As consciousness is intrinsic to a system, IIT considers how a system generates integrated information about itself. Thus, in the example of the two neurons, Φ would assess how the states of the neurons (the “system state”) at time *t* constrains their states at some other time, namely *t* + 1 and *t* - 1. In this manner, Φ reflects how much a system state constrains the possible states the system may take at some other time in the past and in the future. Constraints to both the past and future are considered in order to distinguish purely feedforward or purely feedback networks as non-integrated. This concept extends to an arbitrarily large system – for example, consider two brains as a single system. While considering the present state of both brains together may allow us to predict their future system states better than chance (information), we might also consider the two brains independently and predict just as well (no integration). In other words, considering the two brains together gives us no more information than just considering one brain at a time, and so there is no integration and thus no Φ across the two brains. At any one time, the system state is associated with a probability distribution of states which it may transition into, and a probability distribution of states from which it may have transitioned from. If the system is reducible into independent parts, then the system, when split into its constituent parts, generates the same probability distributions. Thus, Φ is calculated as the “distance” between these sets of probability distributions.

<MIP>Considering the two brains independently is not the only possible way of splitting the “whole” system, however. While there is only one way to split a system of two neurons (by ignoring the connection between them), the number of ways to partition a system grows exponentially with the number of system elements. Other simple possible splits for the two-brain scenario might be to consider the two left hemispheres as independent from the two right hemispheres, or to consider three hemispheres as independent from the remaining hemisphere. Out of all possible partitioning schemes, the one which generates a probability distribution most similar to the whole system is used to calculate Φ. This is akin to identifying the information which is not integrated across the whole system, or reducing the system as much as possible into causally independent parts. The partitioning scheme which yields such a partition is the minimum information partition (MIP). It provides the system in its most reduced state.

<Discussion of studies involving information/integration>While Φ in its latest revision has yet to be calculated in a biological system, the concepts of information and integration seem to be indicative of conscious level. While TMS in wakeful participants triggers responses in multiple cortical areas, these responses become localised and more stereotypical (and shorter) under anaesthesia, indicating reduced effective connectivity {Ferrarelli, 2010 #60}. The same localisation and stereotyping of responses is also observed in NREM sleep, when compared to REM sleep {Massimini, 2010 #61}, suggesting that the loss distinct activity and integration across the brain is a non-anaesthetic specific result of LOC. Furthermore {Sarasso, 2014 #48}’s perturbational complexity index (PCI), which assesses information and integration in EEG responses to a TMS pulse to the thalamocortical system, suggests that vegetative state, minimally conscious, and locked-in syndrome patients have increasing amounts of information and integration. The measure additionally identifies if anaesthetised patients are dreaming {Sarasso, 2015 #63}. Though supportive of information and integration, these studies rely perturbing the brain (e.g. through a TMS pulse) – what is measured is how the perturbation spreads throughout the brain. IIT, however, in principle does not rely on perturbations. Furthermore, assessment of TMS responses assumes a biological system with a brain organised similarly to that of the human’s. Unconsciousness from seizures – loss of information via synchrony; breakdown of cortical effective connectivity during LOC

<Discussion of simulation studies – not needed, maybe have in discussion>

<Discussion of limitations>Despite these promising results, Φ has significant drawbacks. Φ is derived only for discrete variables, but brain recordings are generally continuous. Consequently, its calculation requires complete knowledge of transition probabilities between system states, which is generally infeasible to estimate, especially from a biological system, as this knowledge is gathered from observing all elements in a system (e.g. neurons in a brain). Φ also suffers from high computational costs – as the number of ways to partition a system grows exponentially with the number of elements in the system, the search for the MIP with which to calculate Φ grows exponentially with the number of elements in the system.

<Title> A derivative measure of integrated information

<Introduction to phi-star>To overcome these limitations, several derivative measures of phi, based on concepts from 2.0, have been proposed {Barrett, 2011 #39}. One of these is Φ\*, or integrated information from the decoding perspective. In IIT 3.0, Φ is computed is calculated as the distance between two “uncertainty distributions” (the state probability distributions before and after partitioning into the MIP). Φ\* can be understood simply as the difference in information generated by the whole system and the information generated by the partitioned system. Though the issue of searching for the MIP remains, phi-star overcomes the limitations. Whether it is correlated with the latest derivation of phi is unknown, however.

<Limitation to phi in general; maybe include in phi-star paragraph>However, the computational complexity of searching for the MIP remains a limiting factor in calculating PHI, especially in a system comprising many elements such as the mammalian brain.

## Testing the IIT in Animal Models

Invertebrates present a potential model with which to apply the principles of IIT and calculate Φ. They are particularly useful due to the relatively small number of neurons making up their brains. This has allowed for the complete mapping of the connectome of some invertebrates, such as … Invertebrates with more complex brains are estimated to have their connectomes mapped within the next x years (for drosophila). Meanwhile, mapping of the rat cortex remains steady, but slow (is there an estimated completion date?). The nematode, for example, has had its entire neural network mapped, while mapping of the rat brain is an expensive, ongoing effort.

Due to their significantly different biology and simpler brains, however, it is not immediately clear. Whether an animal is conscious or not however is a question which is posable to any vertebrate, such the mouse, dog, vegetative state patients, or even the people you see each day. Just as we extend the trait of being conscious to animals similar to us, we can make a guess as to whether animals dissimilar to ourselves are conscious based on their neural functioning and behavioural repertoire.

While invertebrates such as the roundworm and nematode exhibit only simple behaviours, other invertebrates display a wide repertoire of behaviours. Some cephalopods, such as the common octopus even display behaviours similar to vertebrates.

Despite being biologically very dissimilar to humans (arguably much simpler), they still exhibit a wide range of behaviours. Furthermore, despite this dissimilarity, brain phenomena observed in vertebrate brains have been observed to some extent also in the fly brain. Don’t need to stick to flies, maybe can talk about other potential models (rat, bee, etc)

Drosophila appears to sleep {Shaw, 2000 #65;Hendricks, 2000 #64} with distinct sleep stages {van Alphen, 2013 #66}.

Other models would also be relevant. As one of the major advantages of IIT is its non-specificity to humans.

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.

## Aims and Hypotheses

Though it has been investigated in simulation studies, the latest formulation of phi provided by IIT 3.0 has not yet been empirically tested as a measure of consciousness 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, and furthermore it is hypothesised that phi will be correlated with phi. 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 Φ\* and its components. It is expected that Φ\* will be correlated with Φ. Furthermore, it is expected that MIP cuts for Φ\* match those for phi.

# **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, set up 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 sharpened fine tungsten wire (0.01 inch diameter, A-M Systems) was inserted into 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 a gas chromatography procedure described by {van Alphen, #66} 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. 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.

### Φ Computation

Data processing for computing Φ was conducted using Python 3.6.0 in MASSIVE (Multi-modal Australian ScienceS Imaging and Visualisation Environment), a high-performance computing facility suited for data processing. To calculate Φ values, I used the PyPhi (0.8.1; Mayner, Marshall, & Marchman, 2016) package for Python 3 to calculate Φ values and their associated MIPs. The mathematical details for calculating Φ are provided in {Oizumi, 2014 #45}. Overall, the calculation of Φ requires a network, its state, and its transition probability matrix (TPM). In the following subsections, I describe how I obtained these inputs. The overall pipeline for computing Φ is presented visually in Figure X.

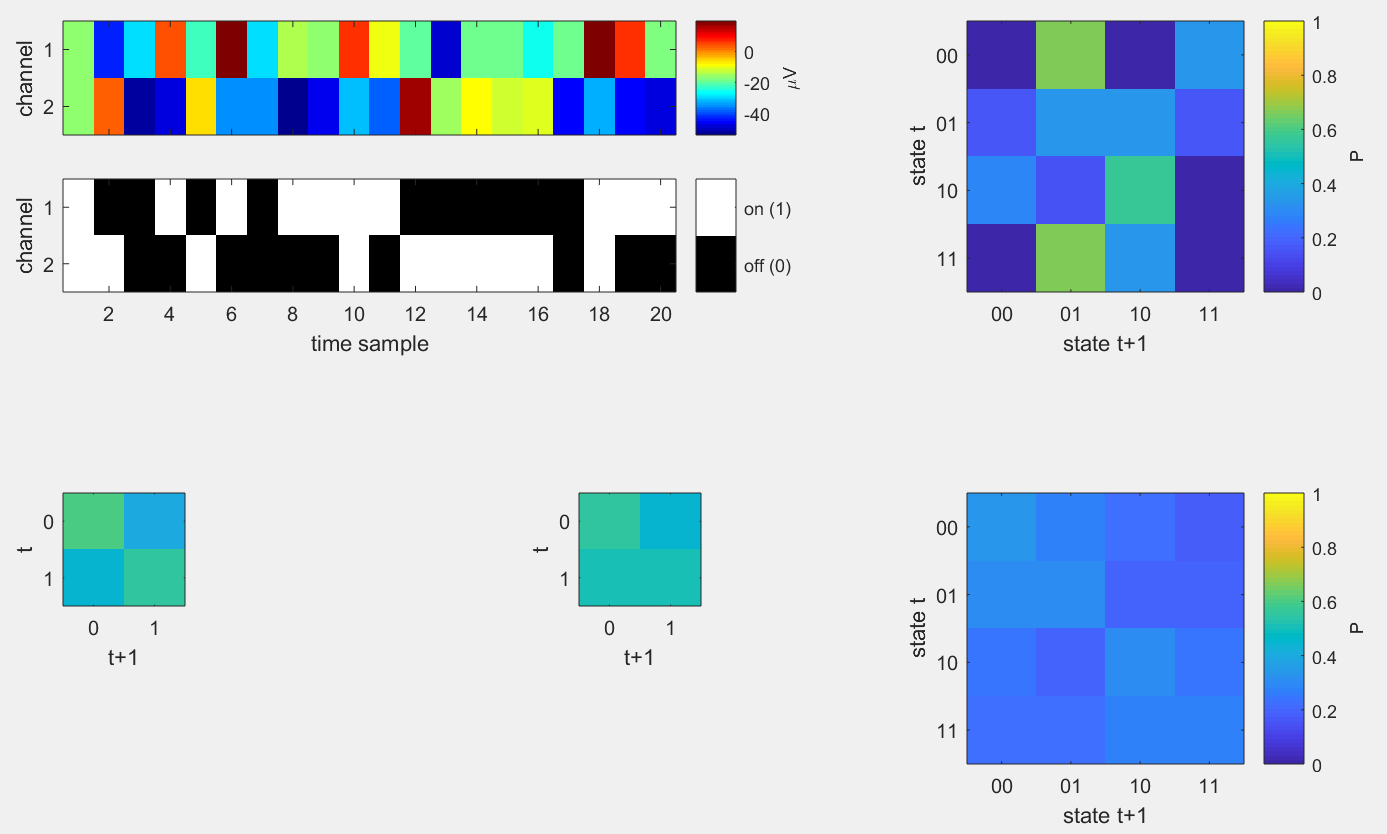


FIGURE X: electrode insertion, rereferenced data, discretised data, tpm

Discretisation. As the latest version of IIT has yet to be extended to continuous variables, discretisation of the continuous LFPs at each channel was necessary. To achieve this, I binarised the recordings of each channel using its median, as taken across samples over all eight trials at a single condition (air or isoflurane). Samples were then replaced with a 1 if greater than the median, and a 0 otherwise. Discretisation in this manner allows us to determine the state of a channel at a given time sample, and thus also the state of a set of channels at a given time sample.

Network Selection. Networks were sets of channels. To avoid arbitrarily selecting channel sets, I selected all combinations of two, three, and four channels, giving 105, 455, and 1365 channel sets respectively. I did not calculate Φ for combinations of more than four channels for the following reasons: (a) as the number of elements being considered increases linearly, the computing time and computing space required to calculate Φ grows exponentially, and (b) as the number of elements being considered increases linearly, the number of possible channel combinations grows super-exponentially (how exactly does nchoosek grow with k?). To accommodate the search for the MIP, all channels within a network were considered to be fully connected (i.e., each channel was considered bidirectionally connected to every other channel); to consider otherwise would limit the possible cuts to make to the system when searching for the MIP. The state of a network at a given time sample is given by the discretised states of its channels (e.g. for channels A = 1, and B = 0, the network state for AB is 10).

Transition Probability Matrix Construction. A transition probability is the probability of a state at time *t* transitioning into another state at time *t* + τ (i.e. the number of times a transition to a specific state occurred divided by the total number of transitions to every state). The transition probability matrix for a network thus holds the transition probabilities of all states at time *t* transitioning into all other states at time *t* + τ: each row of the matrix gives the probability distribution of a given state transitioning into every other state (the “effect repertoire”), while conversely each column gives the probability distribution of states which could have preceded a given state (the “cause repertoire”). Conceptually, Φ is assessed by comparing the cause and effect repertoires of the whole network to the cause and effect repertoires generated by a system split into independent parts.

To ensure adequate sampling to obtain accurate transition probabilities, TPMs were built per condition using all samples across all trials. IIT’s exclusion postulate advocates for calculation at the optimal temporal resolution. However, it is unclear what this resolution is. I thus calculated transition probabilities at three τ levels: 4, 8, and 16 ms.

Collapsing across samples. As Φ is calculated for every state, As these input requirements are met at each time sample, each time sample is associated with a Φ value. To determine the overall Φ for a trial of 2250 samples, I took the average Φ value across samples, weighted by the number of occurrences of each state within the trial. Finally, I averaged Φ across trials.

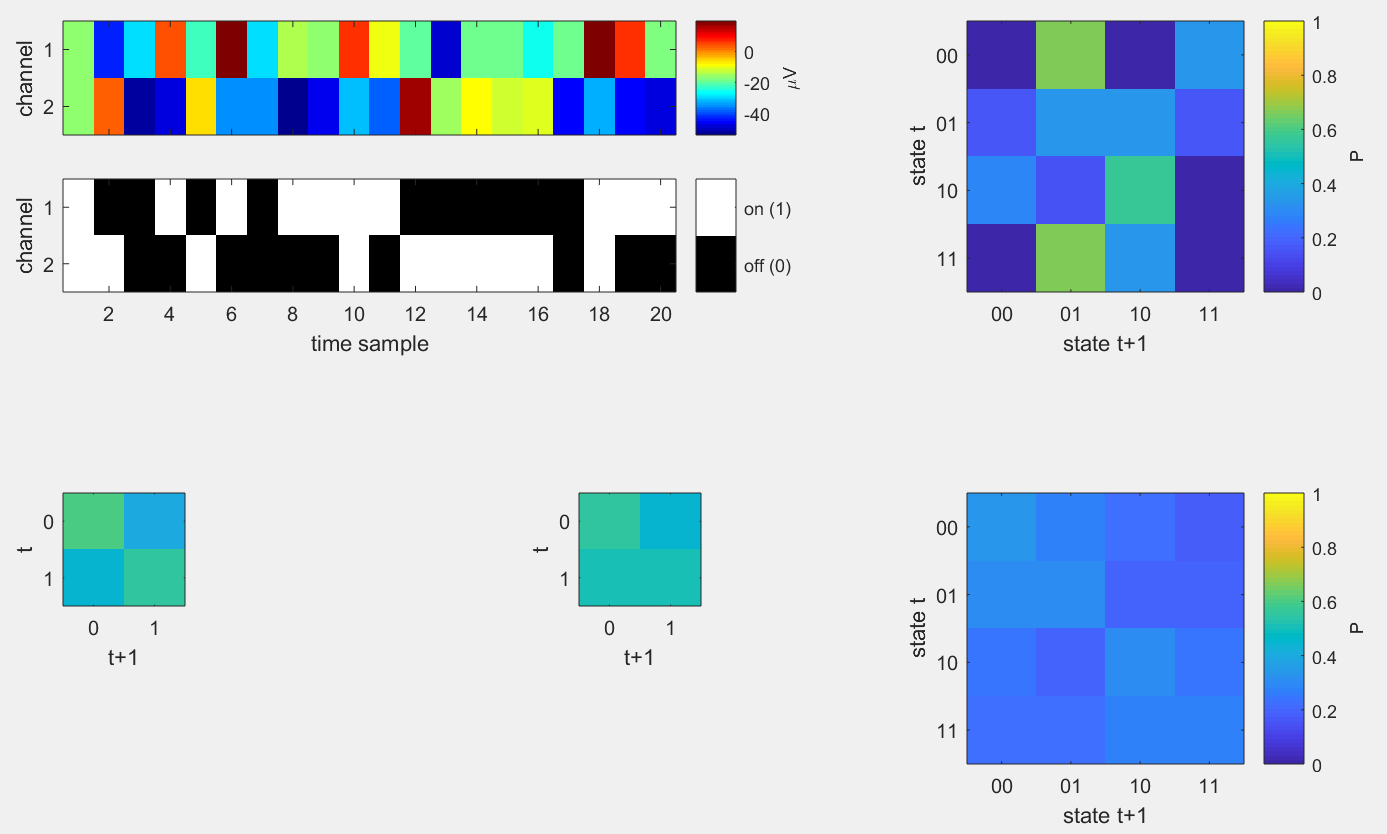


FIGURE X: electrode insertion, rereferenced data, discretised data, tpm

### Φ\* Computation

Data processing for computing Φ\* was conducted using MATLAB R2016a in MASSIVE. While Φ is calculated from a network state and TPM, Φ\* is calculated given a set of continuous signals. Thus, preprocessed LFPs were not discretised as in the computation of Φ. Network selection followed the same paradigm as Φ computation.

Covariances. Φ\* is calculated using time series data. Instead of a TPM, information is assessed by calculating the covariances of the system’s time series with its time series with lag τ. Thus, the covariances across the signals of each channel were calculated.

Φ\* calculation. I used a toolbox which implemented Φ calculation in a previous project {<https://github.com/amhaun01/phipattern>, Haun}. The mathematical details for calculating Φ\* are provided in {Oizumi, 2016 #46}. To find the MIP, I calculated Φ\* across all possible partitioning schemes. The partition which produced the least Φ\* after normalisation {Haun, 2016 #42} was taken as the MIP, and its unnormalised Φ\* as the Φ\* value for the system. As with Φ, I averaged Φ\* across trials.

### Data Analysis

Statistical analyses were conducted using MATLAB R2017a and MATLAB R2015b (for simulated likelihood ratio tests).

Air versus isoflurane. I employed a linear mixed effects model (LME) as an omnibus test for effects of condition, network size, and τ lag on Φ values. Thus, Φ was modelled as dependent on the fixed effects of condition, network size, and τ lag. To account for networks being nested within flies, I included random intercepts for fly and the interaction between fly and network. Fixed effects were tested using simulated likelihood ratio tests (N = 1000) between the full model and a null model with the effect of interest removed.

I assessed change in feedback influence across conditions by taking the assessing MIP cuts in networks with two elements. Feedback was defined as an influence from a central channel to a periphery channel. Only systems with two elements were considered as previous feedback analysis on the same dataset by {Cohen, #2} was bivariate. Furthermore, in systems of two elements, the MIP contains only one cut (either periphery to central, or central to periphery). The same channel grouping scheme as used by {Cohen, #2} was used: channels 2-7 were grouped as peripheral, and channels 10-15 were grouped as central. All other channels were ignored. MIP cuts from a centre channel to a peripheral channel were considered as feedback cuts. As each sample gives a network state and a corresponding MIP, I took the portion of samples within a trial as a measure of feedback influence. After averaging portions across trials and then channel sets, I used a two-way ANOVA to find effects of condition and lag on feedback influence.

Φ versus Φ\*. As with Φ, I employed an LME to test for effects of condition, network size, and τ lag on Φ\* values to account again for the nesting of networks within flies. Once again, trial averaged values were log transformed to address heteroscedasticity, and simulated likelihood ratio tests (N=1000) comparing the full model with null models were used to test for fixed effects.

To assess the convergent validity of Φ\* to Φ, I calculated Pearson correlation coefficients between Φ and Φ\* within each fly and at each network size and lag. Correlations in this manner were calculated across channel sets. I also assessed MIP equivalence between Φ and Φ\*. To do this, I ignored the directionality of Φ MIP cuts, as Φ\* cuts are non-directional. Additionally, as Φ MIP cuts only bipartition the system, trials in which Φ\* was associated with a MIP which was not a bipartition (i.e. consisted of more than two groups) were excluded from the analysis. Because MIP cuts are always equal for two channels (as there is only one way to partition a system of two elements, ignoring directionality), I limited this analysis to channel sets of only three and four channels. MIPs were considered equal if each subgroup in the partition consisted of the same channels. As each trial results in multiple Φ MIPs (with each state being associated with a MIP) but only one Φ\* MIP, I used the portion of samples within a trial for which the sample MIP matches the Φ\* MIP for the whole trial as an indicator of MIP equality. After averaging match portions across trials, and channel sets, I used two-way ANOVAs at each network size to test for effects of lag and condition on match portions.

## Results

### Integrated information is reduced under isoflurane

LME analysis was conducted to assess how Φ behaves with network size, τ lag, and with condition. Due to heavy positive skew of Φ values, trial averaged Φ values were log transformed before fitting the model to address heteroscedasticity. Number of channels, lag, and condition were all significant effects. Φ increased with number of channels (χ2(1) = 1.18 × 105, *p* = .001), and decreased with longer lags (χ2(1) = 4.11 × 104, *p* = .001). Importantly, Φ was decreased in the isoflurane condition (χ2(1) = 3.36 × 104, *p* = .001).

Figure X shows averaged Φ values at each lag for every candidate network at one lag level. Using post-hoc Wilcoxon signed-rank tests on untransformed values at each channel set (with FDR corrected *p* < .05) revealed increasing percentage of sets with a significant reduction in Φ under isoflurane with the number of channels being considered.

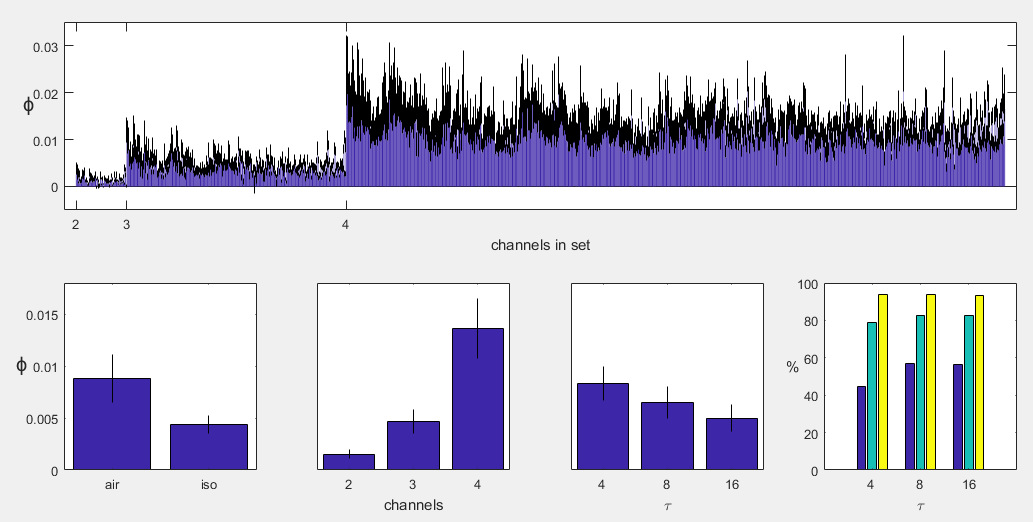


FIGURE X: Average change in Φ values (air – iso) across flies (*N* = 13) (a) Change in Φ for all channel sets at τ = 4 ms. Error bars represent standard error across flies. (b-d) Φ values after averaging across all channel sets within flies, and (b) across set sizes and lags (effect of condition), (c) conditions and lags (effect of set size), and (d) conditions and set sizes (effect of lag). All main effects were significant. (e) Percentage of channel sets with a significant decrease Φ, at each set size and lag. Colours dark blue, cyan, and yellow correspond to sets of 2, 3, and 4 channels respectively.

These changes in Φ were not accompanied by a change in the proportion of feedback MIP cuts in a trial. A two-way ANOVA found no effects of lag (*F*(2, 72) = 1.64, *p* = 0.20) or condition (*F*(1, 72) = 0.05, *p* = 0.82) on the proportion of feedback cuts after averaging across trials and channel sets. The proportions of feedback cuts in either condition were also not significantly different from chance (see Figure X, *p* > .05 for both conditions at all lags).

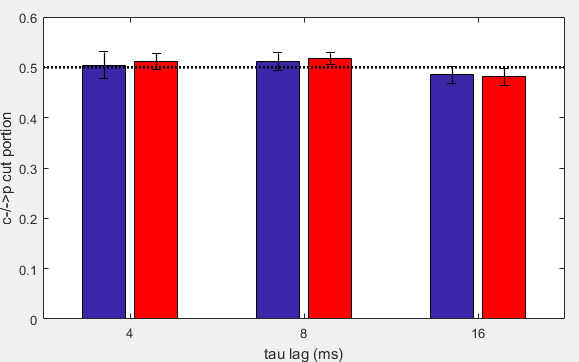


Figure X: Proportion of feedback cuts in the MIP for channel sets of size two. Blue bars indicate the proportion in the air condition, and red bars indicate the proportion in the isoflurane condition.

### Φ\* is correlated with Φ

As with Φ, the behaviour of Φ\* with respect to network size, τ lag, and condition was assessed using LME analysis. Once again due to heavy positive skew, trial averaged Φ\* values were log transformed before fitting the model to address heteroscedasticity. As with Φ, Φ\* increased with the number of channels considered, (χ2(1) = 1.19 × 104, *p* = .001), decreased with longer lags, (χ2(1) = 481.65, *p* = .001), and was reduced under isoflurane (χ2(1) = 1.67 × 104, *p* = .001). The effects of number of channels and lag were not as clear as for Φ after averaging across all channel sets (see Figure X).

Figure X shows averaged Φ\* values at each lag for every candidate network. Post-hoc Wilcoxon signed-rank tests (with FDR corrected *p* < .05) once again showed increasing percentage of sets with a significant decrease in Φ\* for increasing number of channels considered, though not as drastically as for Φ. Unlike Φ, the portion of channel sets with significantly reduced Φ\* was noticeably larger for a lag of 4 ms. Overall, Φ\* was decreased in a smaller portion of channel sets, when compared to Φ (paired t-test).

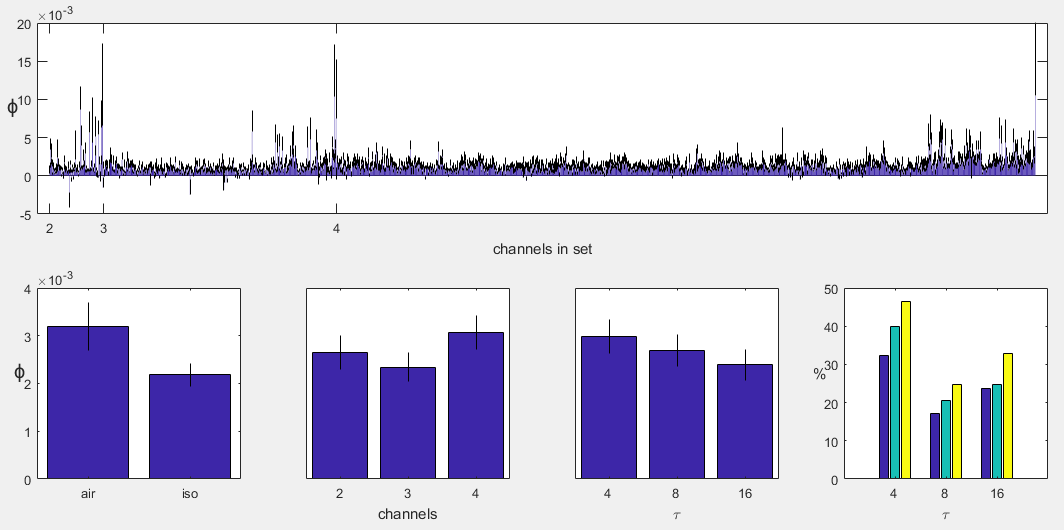


FIGURE X: Average change in Φ\* values (air – iso) across flies (*N* = 13) (a) Change in Φ\* for all channel sets at τ = 4 ms. (b-d) Φ\* values after averaging across all channel sets within flies, and (b) across set sizes and lags (effect of condition), (c) conditions and lags (effect of set size), and (d) conditions and set sizes (effect of lag). All main effects were significant. (e) Percentage of channel sets with a significant decrease Φ\*, at each set size and lag. Colours dark blue, cyan, and yellow correspond to sets of 2, 3, and 4 channels respectively.

Figure X shows the relationship between Φ and Φ\* for one fly at each lag parameter. Correlation coefficients across flies were spurious for sets of two channels. For sets of three and four channels, however, correlations were stronger for shorter lags.

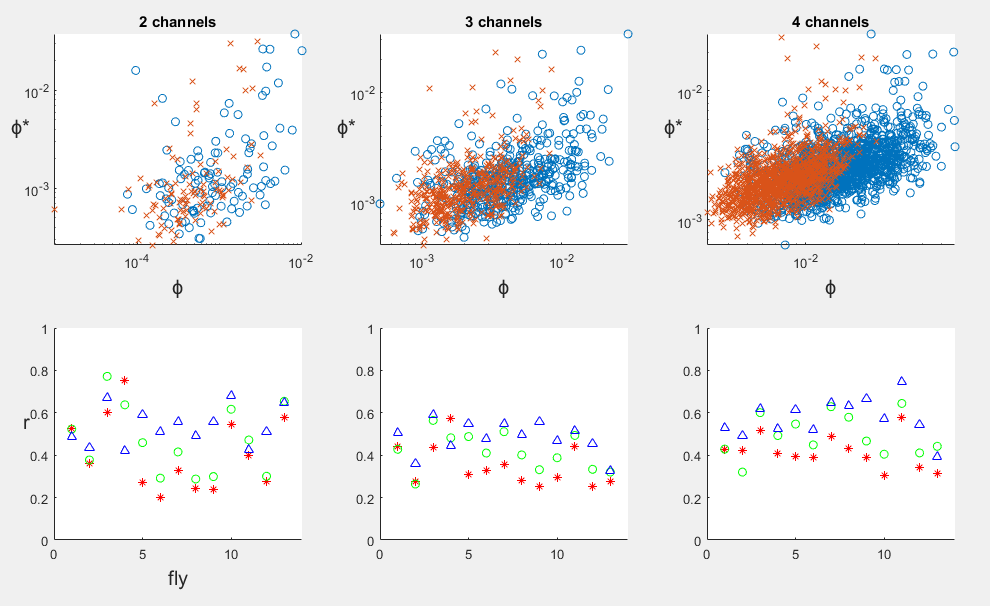


FIGURE X: Relationship between Φ and Φ\*. (a-c) Relationship between Φ and Φ\* for one fly at 4 ms, at each network size. Each point is the trial-averaged Φ and Φ\* value for a channel set. Orange x’s are during isoflurane, and blue circles are during the air condition. (d-f) Φ\* correlation coefficients to Φ for each fly at all lag parameters (red, green, and blue are 4, 8, and 16 ms respectively) across both conditions, at each network size.

Φ MIPs in each trial matched more slightly better than chance (shown in Figure x), but this was significant only when three channels were considered. Two-way ANOVAs at each channel set size revealed no significant effects of either condition (*F*(2, 72) = 0.74, *p* = 0.48, and *F*(2, 72) = 0.73, *p* = 0.49 for three and four channels respectively) or lag (*F*(1, 72) = 0.25, *p* = 0.62, and *F*(1, 72) = 1.02, *p* = 0.32, for three and four channels respectively) on the proportion of Φ/ Φ\* MIP matches per trial.

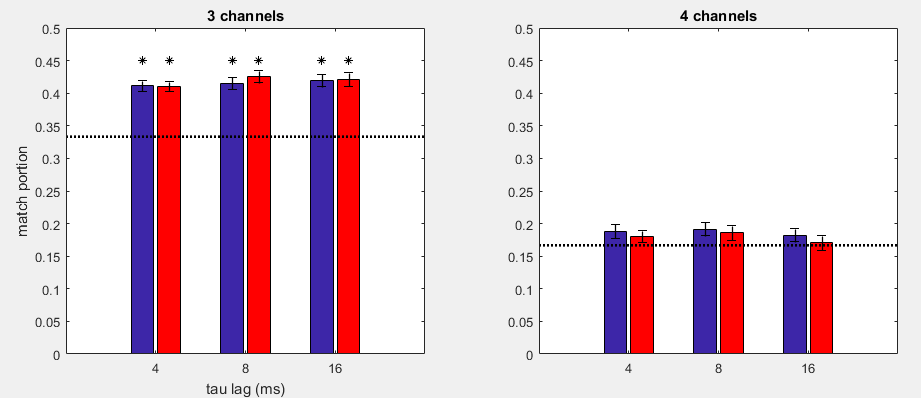


FIGURE X: The proportion of Φ MIP cuts in a trial matching that of the Φ\* cut for the trial, averaged across trials, channel sets, and flies. Blue bars indicate proportions in the air condition, and red bars indicate proportions in the isoflurane condition. Error bars represent standard error across flies. The dotted lines represent expected match proportions due to chance. Asterisks indicate match proportions significantly different from chance, after Bonferroni corrections for multiple comparisons.

## CHAPTER 3: DISCUSSION AND CONCLUSION

The present project aimed to investigate integrated information Φ as a potential measure of conscious level. To achieve this, I calculated both Φ and Φ\* across sets of recordings taken from the fly brain, during and without the administration of isoflurane anaesthesia. To my knowledge, this research is the first to calculate the latest derivation of Φ (from IIT 3.0) in a biological system, and to compare this derivation in varying conscious levels. This chapter is split into three main sections. In the first, I discuss my findings regarding the latest derivation of integrated information, Φ. In the second, I discuss my findings regarding the potentially more practical version of integrated information, Φ\*. Finally, I discuss implications and future directions?

### Φ decreases during isoflurane anaesthesia

Recent literature has associated loss of consciousness with both loss of information and loss of integration. Contributing to such literature are studies which involve patients with …, which utilise newly developed measures based on the concepts proposed by IIT. Less direct links to role of integration include reduction in feedback interactions during anaesthesia. Finally, IIT itself predicts that its measure should be indicative of the level of consciousness in a system. This led to the hypothesis that Φ would be reduced in a biological brain during anaesthesia. Our findings support this hypothesis; Φ calculated from recordings taken during anaesthesia were clearly reduced compared to Φ calculated from recordings taken during wakefulness. This result adds the two known studies which have compared Φ between levels of consciousness in a biological system {Chang Frontiers; Lee Conc & Cog}, both of which used a previous derivation of integrated information, and reported reduced Φ during reduced conscious levels <but did they look at differing network sizes?>.

In addition to reduced Φ with reduced conscious level, I also found increases in Φ with number of channels considered. While the behaviour of Φ, with regards to number of system elements being considered, in biological systems hasn’t been investigated before now, this result is consistent with simulation studies {}. This result also intuitive(?) with regards to IIT. Φ as proposed by IIT is purposed to also reflect a conscious experience, and its magnitude can be and has been <actually interpreted as ‘capacity to integrate’? – find paper> interpreted as a measure of the quantity of possible experiences a conscious system may have. For example, a system consisting of two binary elements can represent more “experiences” (i.e. take more states) than a system of only one binary element. While the present project was limited to systems of up to four elements, it is likely that Φ will continue to increase as more elements are included, so long as they remain integrated.

A key question with regards to Φ is the spatiotemporal resolution at which to calculate it. IIT’s principle of exclusion posits that consciousness cannot be superimposed. Thus, though we can in principle calculate Φ in a system at two different time scales, IIT states that this cannot be interpreted as the one system multiple consciousness, each running in its own timescale. Instead, IIT proposes that the system’s consciousness exists only at the scale at which Φ is maximal. While not a primary aim of this project, my results indicate that maximal Φ may be attained at a more granular timescale. <consistent with any simulation studies?> Whether this extends to larger networks however is unclear. Long range interactions in the x brain occur in the order of x {}. <reaction time / perception resolution of flies>. Thus, it is possible that as more elements are considered, the optimal resolution at which Φ reaches a maximum may increase. Consequently, the present finding of greater Φ at more granular time resolutions should be interpreted with caution.

A recent paper {Cohen 2017} which applied frequency domain Granger causality analysis to the same data set used in the present project identified a reduction in feedback influences during isoflurane anaesthesia (over long timescales; low frequency – <~10 Hz). Furthermore, the general literature has associated reduced conscious level with reduced feedback interactions. <link MIP cuts to GC>.Thus, it was anticipated that MIP cuts from central channels to peripheral channels would be more likely during anaesthesia, especially at the least granular time resolution. This hypothesis was not supported, however, with the likelihood of having a feedback cut as the MIP cut sitting squarely at chance in both the air and isoflurane conditions, at all time resolutions. This suggests that integration as assessed by Φ is not captured through the more standard measure of Granger causality <should probs introduce GC somewhere>. Furthermore, it indicates that <adds to? Is there literature on matching frequency domain analyses to time domain analyses?> frequency domain analyses may not necessarily line up with time domain analyses at corresponding/similar timescales. The lack of replication of reduced feedback should be interpreted carefully however. As a TPM holds the transition probabilities of each state to each other state, it indirectly holds the interactions among system elements. Thus, it is plausible that loss of feedback influences is captured in the differences between TPMs. Consequently, comparing MIP cuts in the contexts of different TPMs, as in the present project, may not reflect such changes in influences among elements.

### Φ\* is correlated with Φ

Recent literature involved in investigating integrated information as a potential measure of consciousness have generally utilised a derivative version of Φ, based on the previous iteration of IIT. The derivative version, Φ\*, is appealing as it addresses the observational limitations of Φ and is analytically calculable under the assumption of Gaussian variables. These strengths make Φ\* more practical to calculate in biological systems, which naturally have many elements. Additionally, the version of IIT upon which it is based is somewhat simpler than the latest, allowing for easier/simpler understanding of integrated information. Despite its popularity however, it has not yet been compared directly to the latest version of Φ. As Φ\* meets the theoretical requirements of the version of IIT upon which it was derived, and the latest version of IIT, while updating certain concepts, is conceptually similar to its previous iteration, I expected Φ\* to behave in a similar manner to Φ, and consequently exhibit convergent validity with Φ, as expressed through correlations between Φ and Φ\*, and above chance likelihood of Φ MIPs matching Φ\* MIPs.

Convergent validity between Φ\* and Φ was partially supported. As anticipated, Φ\* behaved similarly to Φ. However, in contrast to Φ, the effects of number of channels considered and lag were not so drastic as to remain obvious after averaging across channel sets {stats – loss of power when averaging observations}. Furthermore, the proportion of channel sets which underwent a significant reduction in Φ\* was markedly less than that in Φ. Finally, while the proportion of channel sets which experienced a significant reduction in Φ under anaesthesia was comparable across timescales, the number of sets which experienced this in Φ\* was notably higher at the most granular timescale. Φ\* thus was generally only moderately correlated with Φ. Interestingly, correlations between Φ and Φ\* were stronger at less granular timescales, despite higher Φ and Φ\* values at more granular scales. … Despite consistent correlations between Φ and Φ\*, the likelihood of MIPs matching was close to chance. Additionally, though correlation strengths increased with increasing timescales, the likelihood of MIPs matching remained consistent.

### Significance and implications (and future directions)

This research is the first to calculate the latest derivation of Φ in a biological system, and the first to compare this latest derivation in varying conscious levels.

Significance – validity for Φ, validity for Φ\* as a simpler measure

Limitation – Φ vs just information or just integration

Future – identify complex

Future – identify conscious contents

Future – improvements in computing Φ

Conclusions

MIP matching – degree of match

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.

Though the search for the MIP hinders the practicality of computing phi, significant progress is being made to overcome this limitation. Already in IIT 3.0, the minimum information bipartition is used (as if any part of the system is independent, then it will be picked up by some bipartition). Another approach to this problem is to approximate the MIP through clustering algorithms {Toker, 2017 #62}.

## CHAPTER 4: REFERENCES