Integrated information Φ 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 4](#_Toc495225028)

[Statement of Contribution 5](#_Toc495225029)

[CHAPTER 1: INTRODUCTION 6](#_Toc495225030)

[The Search for the Neural Substrate of Consciousness 6](#_Toc495225031)

[The Integrated Information Theory of Consciousness 8](#_Toc495225032)

[Practicality of Φ as a Measure of Conscious Level 12](#_Toc495225033)

[Testing the IIT in the Fly Model 14](#_Toc495225034)

[Aims and Hypotheses 15](#_Toc495225035)

[CHAPTER 2: EXPERIMENTAL METHODS AND RESULTS 16](#_Toc495225036)

[Method 16](#_Toc495225037)

[Experimental Procedure 16](#_Toc495225038)

[Animal preparation. 16](#_Toc495225039)

[Electrode probe insertion. 16](#_Toc495225040)

[Isoflurane delivery. 17](#_Toc495225041)

[Experimental protocol. 17](#_Toc495225042)

[Local field potential preprocessing. 17](#_Toc495225043)

[Φ Computation 18](#_Toc495225044)

[Discretisation. 19](#_Toc495225045)

[Network Selection. 19](#_Toc495225046)

[Transition Probability Matrix Construction. 19](#_Toc495225047)

[Collapsing across samples. 21](#_Toc495225048)

[Φ\* Computation 21](#_Toc495225049)

[Covariances. 22](#_Toc495225050)

[MIP search. 22](#_Toc495225051)

[Data Analysis 23](#_Toc495225052)

[Air versus isoflurane. 23](#_Toc495225053)

[Φ\* versus Φ. 24](#_Toc495225054)

[Results 25](#_Toc495225055)

[Air versus isoflurane 25](#_Toc495225056)

[Φ\* versus Φ 27](#_Toc495225057)

[CHAPTER 3: DISCUSSION AND CONCLUSION 31](#_Toc495225058)

[Φ decreases during isoflurane anaesthesia 31](#_Toc495225059)

[Φ\* is correlated with Φ 33](#_Toc495225060)

[Significance and implications (and future directions) 34](#_Toc495225061)

[CHAPTER 4: REFERENCES 36](#_Toc495225062)

# Abstract

The integrated information theory of consciousness (IIT) 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. Thus, this project aimed to (a) assess the construct validity and behaviour of Φ as a measure of conscious level in a real biological system, and (b) compare it to a potentially more practical derivation of integrated information, Φ\*. To achieve these aims, recordings from the fly brain were obtained during and without administration of isoflurane anaesthesia. I calculated both Φ and Φ\* across increasingly large sets of these recordings, at varying timescales. Both Φ and Φ\* were significantly reduced during anaesthesia, increased with number of recordings, and decreased with larger timescales. Overall, the two quantities were moderately correlated. In addition to supporting IIT’s prediction of reduced Φ with reduced conscious level, these results demonstrate that Φ grows rapidly with the number of elements in a biological system and is maximal at smaller timescales, and validates Φ\* as a practical derivative measure. (200 word limit)

# Statement of Contribution

The present project was produced by the author, in collaboration with Naotsugu Tsuchiya of the Tsuchiya Laboratory, Monash Neuroscience of Consciousness (MONOC) at Monash University. With guidance from Naotsugu Tsuchiya, the author formulated the aims and hypotheses of this project. Data used in this project was provided by Dror Cohen, who previously had collected and pre-processed the data for his PhD thesis.

The author developed the Python, MATLAB, and Slurm scripts and functions to utilise publicly available toolboxes in calculating Φ and Φ\* from the data, in the Multi-modal Australian ScienceS Imaging and Visualisation Environment (MASSIVE). On average, calculation of these measures across all parameters costed approximately 64 hours of continuous computation time per fly. The author also developed the MATLAB scripts and functions with which to analyse and generate visualisations of the resultant data.

The author prepared the final paper, which was critiqued by Dror Cohen and Naotsugu Tsuchiya. The written material presented in this document is the author’s own work, and feedback was obtained once each from Naotsugu Tsuchiya and Dror Cohen.

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

Date: … 2017

# **CHAPTER 1: INTRODUCTION**

We all have an idea of what consciousness is – it is that which is lost when sleep, and re-emerges when we wake or dream. This is a distinct phenomenon from others which may also be referred to as “consciousness”, such as such as self-awareness {Morin, 2006 #84}, high-order thoughts {Edelman, 2003 #85}, or reportable access to one’s own experience {Block, 1995 #83}. Though consciousness is a private experience, limited to ourselves, we generally extend the notion of consciousness to other animals – seemingly purposeful behaviour is sufficient in most cases for us to ascribe consciousness to an entity. If unsure, we might stimulate the being and determine a level of consciousness based on its reaction. Determining if a being is conscious isn’t so straightforward, however. For example, conscious experience is possible without overt behavioural responsiveness, such as when dreaming or, in rare cases, when under anaesthesia {Liu, 1991 #67;Sebel, 2004 #68}. Furthermore, <physiological example>. In such cases, it becomes clear that the general “signs” of consciousness are not truly indicative of consciousness, whether they be behavioural signs {Guedel, 1937 #70} or physiological signs {Rani, 2012 #69}. Thus, a key goal in neuroscientific research is to identify the necessary conditions for consciousness - to find how consciousness arises from neural activity in the brain.

## The Search for the Neural Substrate of Consciousness

In an attempt to understand how consciousness arises from physical interactions, neuroscientific research has largely focussed on finding neural correlates of consciousness (NCC; {Koch, 2016 #12}). Within this approach, researchers have traditionally focussed on identifying the NCC of the levels of consciousness (level NCC; the minimally sufficient conditions to achieve some level of consciousness, such as wakefulness over dreamless sleep), or the NCC of specific contents of consciousness (content NCC; the minimally sufficient conditions to achieve a specific conscious percept, such as that of a face or the colour red). The search for NCCs has led to the observation of numerous specific neural interactions which may form part of the NCC. For example, synchronous activation among neurons {Engel, 2001 #18}, 40 Hz oscillations in the cerebral cortex {Llinás, 1993 #58}, and feedback interactions {Lamme, 2010 #30} have been proposed as parts of level NCC, and activations of cortical areas such as the fusiform face area (FFA) during face perception {Pierce, 2001 #57} have been proposed as parts of content NCC. This reductionist approach to understanding consciousness however has significant drawbacks.

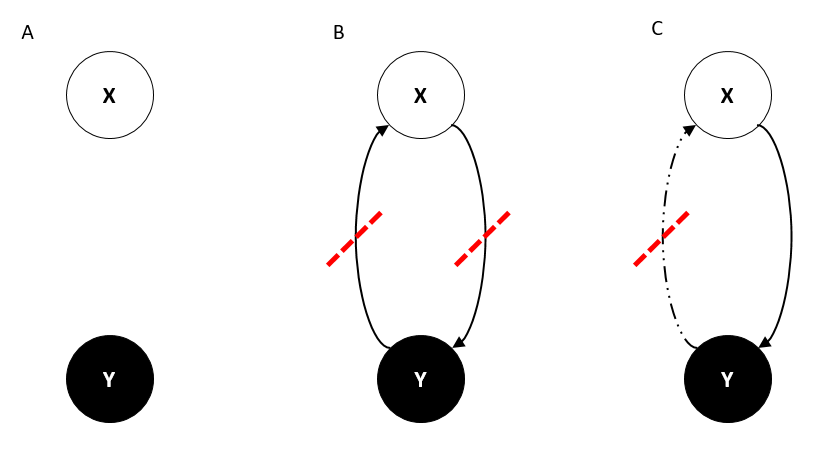
The first drawback is that observed NCC cannot be taken as reliable indicators of consciousness. For example, synchronous activity and feedback interactions both occur in the cerebellum {Person, 2012 #19;Witter, 2016 #29}, which likely does not contribute to consciousness {Yu, 2015 #20}, while the FFA is activated during perception of non-face stimuli {Gauthier, 1999 #71;Gauthier, 2000 #72}. The second drawback is the lack of generalisability of proposed NCCs to systems other than those in which they were observed. For example, the exact brain region activated during perception of human faces varies among humans, non-human primates, sheep, and dogs {Cuaya, 2016 #91}. Furthermore, it is clear that biological NCCs cannot be used to assess consciousness in an artificial system such as a computer. Thus, though NCCs may be informative as to the kinds of structures and interactions which might support consciousness, they ultimately do not explain how these give rise to consciousness. This explanatory gap between the physical substrate of consciousness and consciousness itself, referred to as the hard problem of consciousness {Chalmers}, gives rise to the need for a principled theory of consciousness.

Given the limitations of NCCs, there is now a stronger emphasis on using a more theoretical approach in tackling the question of how consciousness arises. Accordingly, a number of theories of consciousness have been proposed, such as the global workspace theory {Baars, 1997 #4;Baars, 2002 #5} and cross-order integration theory {Kriegel, 2007 #59}. While a common theme among these theories is the notion of integration across parts, they generally leave the concept ill-defined and un-operationalised. For example, the global workspace theory advocates that integration across distant brain regions gives rise to consciousness. While studies linking conscious perception of a stimulus to functional connectivity across distal brain regions {} are taken to support this idea, the theory provides no explanations as to how such regions become integrated. Meanwhile, the cross-order integration theory proposes that consciousness arises from integration between a first-order representation of an external stimulus and a second-order representation of that first-order representation, but doesn’t describe possible ways in which this might be achieved. Consequently, though they provide some testable predictions as to what kinds of large scale biological structures and functional interactions might give rise to consciousness, they ultimately fail in proposing a physical substrate for consciousness. As a result of this, they also fail to provide any measure of the consciousness they aim to explain.

## The Integrated Information Theory of Consciousness

The integrated information theory of consciousness (IIT; {Oizumi, 2014 #45}) stands out from competing theories by taking a different approach towards finding the physical substrate of consciousness. Instead of building a theory from observed neural activities, IIT identifies fundamental aspects of consciousness, and from these it reasons the necessary mechanisms for it. The fundamental properties of consciousness IIT identifies are as follows: (a) intrinsic existence: an experience exists intrinsically to a conscious system, but not for external observers (in other words, observing a conscious system does not give you the experience that the system is having); (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 the conscious system could instead be having (for example, by reading this thesis you are consequently not experiencing all the other experiences you could possibly having instead, such as watching a movie, cooking dinner, or reading a more interesting thesis); (d) integration: an experience exists as a single whole which cannot be broken up into independent parts (for example, the experience of a red ball does not reduce to two separate, independent experiences of redness and of a ball); and (e) exclusion: an experience cannot be superposed with other experiences – they either merge into a single experience or they form separate consciousnesses (this precludes a single system from having multiple consciousness, for example at different timescales). From these fundamental properties of consciousness, IIT derives a set of physical properties from which consciousness arises (the full derivation is described in {Oizumi 2014}. These physical properties lead to a measure, integrated information Φ, whose magnitude is purported to reflect the level of consciousness in a network. As Φ is derived from fundamental principles, rather than based directly on observed neural activity, its applicability is not limited to humans, vertebrates, or even biological systems. Thus, it overcomes the previous limitations of the search for the NCC.

IIT proposes that a conscious network specifies its own causes and effects. That is, in a conscious network, the network state (i.e. the combination of the states of all elements in the network, for example the state of all neurons in the brain) should be informative as to the possible previous network states *and* the possible future network states. Φ is derived to reflect the degree to which a network achieves this. In the next three paragraphs, I attempt to provide a basic conceptual explanation of Φ using simple networks consisting of two neurons (Figure X). For details, its full mathematical derivation, which is unnecessarily complicated for this thesis, is provided in {Oizumi, 2014 #45}.



*Figure X*. (include probability distributions?) Assessment of Φ in networks of two neurons. Each neuron, X and Y, takes one of two states: firing (white; 1) or not firing (black; 0). Of four possible network states (00, 01, 10, and 11), the current network state XY at some time *t* is 10. (a) If nothing is known about the interactions between X and Y, then at time *t* + 1 XY could assume any one of the four possible network states. (b) If X at *t* + 1 assumes Y’s state at *t* and vice versa for Y and X (black arrows), then the network state at *t* + 1 must be 01. The dotted red lines indicate potential MIP cuts – taking the left cut and ignoring Y’s influence on X increases uncertainty as to the state of X at *t* + 1 (possible states are 11 and 01), while taking the right cut and ignoring X’s influence on Y gives uncertainty to the state of Y at *t* + 1 (possible states are 00 and 01). (c) Only Y at *t* + 1 copies X’s state at *t*. The broken black line indicates that the state of X does not depend on the state of Y. Thus, at ­*t* + 1, possible states are 01 and 11. Ignoring X’s independence from Y does not change the possible states, but ignoring X’s influence on Y does (if ignored, all four network states are possible at *t* + 1). Note how in this case the system is purely feedforward from X to Y.

The magnitude of Φ may be understood from the principles of *information* and *integration*. *Information* (the I of Φ) refers to reduction in uncertainty. In our first example network, given no knowledge about the interactions between the neurons, there are four possible network states at any given timepoint *t*. However, if we consider that each neuron at time *t* takes the state the other neuron took at time *t* – 1 (essentially copying each other at each timestep), and the state of the network at timepoint *t*, we would be able to deduce the only possible state of the network at *t* + 1. This reduction from four potential states (high uncertainty) to one (low uncertainty) is information. Depending on the precise interaction between the neurons, this reduction could be less, resulting in less information. The same principal is applied for time *t* – 1. Thus, the “information” of Φ represents the extent to which the state of a network constrains its possible past states (causes) and possible future states (effects).

Integration (the O of Φ) assesses the extent to which a network is irreducible to subnetworks which are connected only through unidirectional interactions. IIT assesses this by partitioning the network and comparing the information generated by the full network to the information generated by the partitioned network. Partitioning is achieved by ignoring causal links from a set of elements in the network to the remaining elements. In our copying neurons example, ignoring either of the links (X to Y, or Y to X), allows us to only increase the set of possible network states from one to two, thus decreasing the information generated by the network. In the third example however, where only Y copies X, cutting the link from X to Y reduces the information generated, while cutting the link from Y to X does not. To reduce the network as much as possible to its constituent parts, we always take the partition which generates information as similarly to the full network (the minimum information partition, or MIP). In this scenario, as the full network gives the same information as when partitioned according to the MIP, the information generated by the whole network is considered not integrated. Thus, the “integrated” in Φ represents the extent to which information generated by a network is lost when a network is partitioned.

Putting information and integration together, Φ can be understood as the extent to which a network state constrains the network’s possible past and future states, and simultaneously the extent to which that constraint is lost when the network is split into smaller networks interacting only through either feedforward or feedback connections. This concept extends to arbitrarily large systems – for example, consider two brains as a single network. Though the network states of both brains at some timepoint together constrain their collective possible network states at some other time, this constraint is likely no more than that of the two brains independently constraining their own network states. 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. Dividing the two-brain network into two separate brains however is not the only potential way of partitioning the network. We may, for example, partition the network into a group of two left hemispheres, and two right hemispheres. Given that brain hemispheres are integrated to some degree {}, information generated from this partition would thus likely be lower than the full network, indicating integration. Thus, the purpose of using the MIP to assess integration is to reduce the network as much as possible to independent parts.

### Integration, Information, and Conscious Level

Information and integration in the brain seem to change with conscious level. For example, while transcranial magnetic stimulation in wakeful participants triggers responses in multiple cortical areas, these responses under anaesthesia become localised and stereotypical, indicating reduced effective connectivity among regions {Ferrarelli, 2010 #60}. The same localisation and stereotyping of responses is observed also dreamless non-rapid eye movement sleep, when compared to rapid eye movement sleep {Massimini, 2010 #61}, suggesting that the loss of distinct activity and integration across the brain is a general result of reduced conscious level. Furthermore, the assessment of information and integration electroencephalographic (EEG) recordings has been demonstrated to identify patients under anaesthesia who are dreaming. Finally, variations of Φ itself has been demonstrated in EEG recordings to decrease in humans during administration of anaesthesia {Lee, 2009 #44}, and during sleep {Chang, 2012 #88}. These Φ studies however are based on an older derivation of Φ {Tononi, 2008 #50}, which only assesses a network’s ability to specify its causes, rather than causes and effects, and calculates information using a different mathematical method {Oizumi, 2016 #46} to the latest version. Meanwhile, the latest derivation of Φ has yet to be calculated in a biological system.

### Practicality of Φ as a Measure of Conscious Level

Though level of consciousness is clearly linked to information and integration in the brain, Φ as a measure has several drawbacks which limits its applicability to biological systems. Firstly, information generated by a network is assessed by finding the transition probabilities of each network state to every other network state. However, as the number of possible network states grows exponentially with the number of network elements (for example, while a network of four binary elements has 16 possible states, a network of five elements has 32 possible states), obtaining complete empirical knowledge of transition probabilities between network states becomes a serious challenge in large networks. Furthermore, while the calculation of information requires assumptions which are not met in continuous variables {Oizumi, 2016 #46}, common mid- to large- scale brain recordings which collate across populations of neurons, such as EEG and local field potentials (LFP) are continuous in nature. Secondly, assessment of the integration among network elements into a single whole depends on identifying the MIP. However, without prior knowledge of a network’s causal connections, the MIP can only be found by assessing every partition of the network. An exhaustive search for the MIP in this manner is computationally expensive, as the number of ways to partition a network grows super-exponentially with the number of elements in the network (for example, while a network of 4 elements can be partitioned into subnetworks in four ways, a network of five elements can be partitioned in 15 ways {Aitken, 1933 #76} {Bell, 1934 #77}{Toker, 2017 #62}). Consequently, the calculation of Φ across a network of 1000 elements is estimated to take several magnitudes longer than the age of the universe {Toker, 2017 #62}.

To overcome these issues, several approaches have been taken. With regards to the massive number of observations needed to determine transition probabilities between a large network’s states, several derivative versions of Φ have been proposed, such as ΦE {Barrett, 2011 #39} and Φ\* {Oizumi, 2016 #46}. ΦE is derived to be applicable to continuous time-series data, and thus overcomes the requirement of determining transition probabilities between discrete network states. Its concepts are extended in Φ\* to meet the theoretical requirements of Φ {Oizumi, 2016 #46}. Versions such as these are however built from a previous version of IIT, and so they do not fully assess a network’s ability to specify both its causes and effects. There currently are no derivative measures based on the latest version of IIT.

Another approach aims to reduce the computational costs of calculating Φ associated with the search for the MIP. Considering bipartitions (and thus the minimum information bipartition, MIB, rather than the MIP) instead of all partitioning schemes significantly reduces the set of partitions to search over, but this search still grows super-exponentially with the number of elements in a network. Limiting the search further to bipartitions which split a network into equal halves still gives exponential growth {Toker, 2016 #78}. Recent approaches to further reducing the search space for the MIP include approximating the MIP from inferring likely partitioning schemes, based on connectivity within a network {Toker, 2017 #62}. Reduction in search space for the MIP or MIB in such a manner significantly reduces search time when compared to repeatedly recalculating Φ for each possible partition.

By both using derivative versions of Φ, including Φ\*, and approximating the MIP, Toker and Sommer (2017) managed to calculate Φ across single unit recordings across the nematode brain. However, as the latest theoretical Φ has not yet been calculated in a biological system, derivative measures such as Φ\* have not been compared to the original measure. Given that derivative versions assess a network’s ability to constrain only its causes, as opposed to both causes and effects, it is unclear as to whether their assessment of integration, as done by searching for the MIP, is equivalent to that of the latest theoretical Φ. Consequently, it is unclear if these derivative versions reflect Φ as measures of conscious level.

### Testing the IIT in the Fly Model

Ultimately, largest drawback to using Φ as a measure of conscious level is the difficulty in applying it to large networks such as a brain. Thus, the question of where to test the measure arises. With this context, the fly model presents as a potential model in which to assess Φ.

The fly brain is potentially much simpler than the mammalian brain. Compared to mammalian brains, it has several orders of magnitudes fewer neurons (~135,000 in the fruit fly brain compared to ~70,000,000 in the mouse brain; {Alivisatos, 2012 #24}{Herculano-Houzel, 2006 #26}). Because of this, mapping of the fly connectome is estimated to reach completion within the next x years {}, much sooner than mapping that of a mammal’s, such as the rat connectome (estimated for completion in y years; {}). In a similar vein, due to having fewer neurons and thus requiring fewer observations to characterise all interactions among neurons, calculating Φ across the fly brain presents a much more feasible goal than doing the same in a mammalian brain.

Despite the relatively small number of neurons in their brain, flies still exhibit a variety of behaviours. This is in contrast to other, even more simpler, potential models. While other invertebrates such as the nematode and roundworm exhibit only simple behaviours, the fly exhibits both short and long term memory {}, and complex learning {}. While these behaviours alone are not sufficient to imply consciousness, as in humans, we apply these to infer consciousness in other mammals {really?}. Furthermore, flies exhibit alterations in activity, suggestive of altered conscious level, such as torpidness similar to a sleep state in mammals. Just as in mammals, such sleep states have distinct states. <Sleep, inactivity during anaesthesia>

In addition to behavioural similarities to mammals, administration of anaesthesia affects the fly in a similar as to mammals. Flies enter a state of reduced activity, and the neural mechanisms through which this is achieved is comparable to mammals. For example … Further supporting the notion of similarity between the fly brain and mammalian brains. Furthermore, feedback interactions have been reported to decrease under anaesthesia.

Finally, one of the major appeals of IIT is its in principle applicability to any network.

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.

## Aims and Hypotheses

To my knowledge, the latest formulation of Φ has not yet been calculated across neural recordings obtained from a biological system. Consequently, it has not yet been empirically assessed as a measure of conscious level, and the ideal timescale at which to calculate it is not yet established. Additionally, though the approximation of Φ, Φ\*, has previously been calculated across biological recordings {}, its convergent validity to the original measure has not been assessed. Thus, this project had two primary aims.

The first aim was to compare Φ in the fly brain during wakefulness and during isoflurane anaesthesia. Within this aim, I hypothesised that Φ would be reduced during anaesthesia. Given that I used a subset of the data analysed in {Cohen, 2017 #93}, who described reduced feedback influences in the fly brain during anaesthesia, I also hypothesised that MIP unidirectional cuts from the centre of the brain to the periphery would be more likely during anaesthesia than cuts from the periphery to the centre.

The second aim was to compare Φ\* to Φ. As the overall concepts between the two versions of Φ are similar, I hypothesised that Φ\* would also be reduced anaesthesia. As Φ\* would ideally be equivalent to Φ, I hypothesised that Φ\* and Φ would be positively correlated. Furthermore, and that the likelihood of Φ\* MIPs matching with Φ MIPs would be greater than chance.

A final aim was to assess how Φ changes with number of elements and timescale. Given that uncertainty increases with number of elements, I hypothesised that Φ would increase with more elements. While the timescale of integration is unclear, <Dror’s coherence> led me to hypothesise that Φ would be greatest at …

# **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 glued dorsally to a tungsten rod. 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 able to walk on an air-supported Styrofoam ball {Paulk 2013}.

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. Probes had an electrode site separation of 25 µm and measured 375 µm from base to tip. As a reference electrode, a sharpened fine tungsten wire 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, and subsequently observing visually evoked potentials {} 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}.

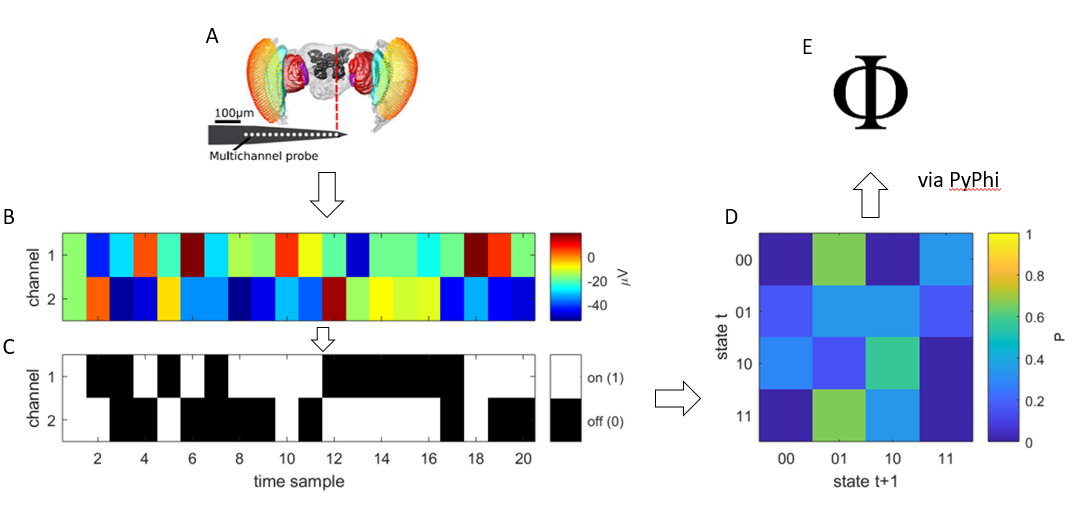
Isoflurane delivery. Isoflurane was delivered from an evaporator onto the fly through a connected rubber hose. The isoflurane was delivered at a constant flow of 2 l/min and continuously vacuumed from the opposite side of the fly. Actual concentration near the fly body was either 0 vol% (air condition) 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}. Hereafter, I use the term “condition” to refer to conscious level as manipulated through isoflurane.

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. These preprocessed data were provided to this project. The following methods describe the procedure directly used in this project to calculate Φ.

### 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, 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 calculating Φ is presented visually in Figure X.

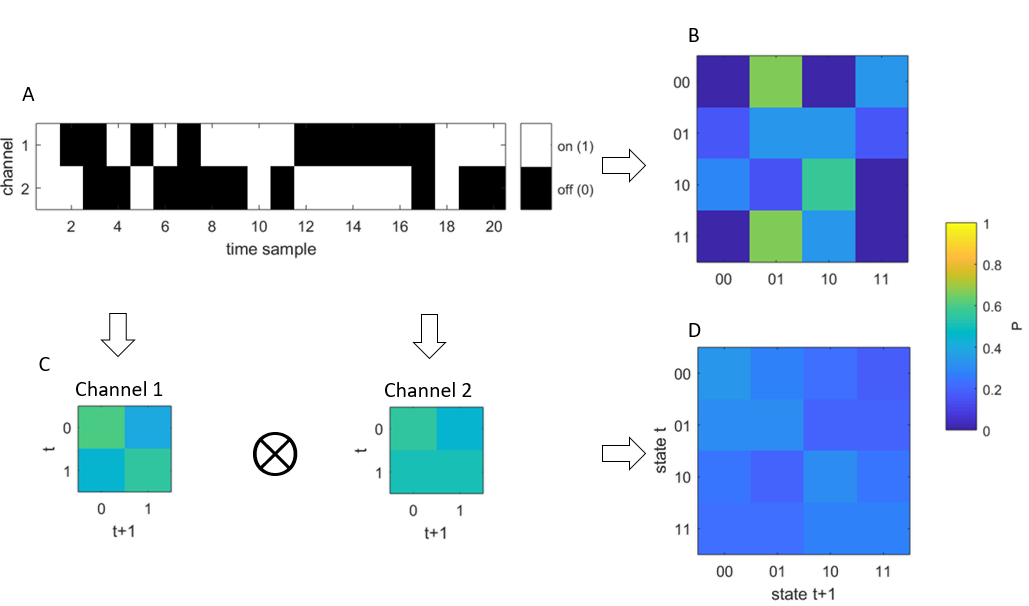


*Figure X.* My processing pipeline for calculating Φ. (a-b) A linear multi-electrode probe records LFPs from the fly brain. Displayed in (b) are 20 samples for a network of two channels, after pre-processing. (c) Samples in each channel are discretised into one of two values. The state of the network at a given sample is given by the states of its channels. For example, the first sample has a state of 11. (d) A TPM is built from discretised samples (for ease of explanation, τ = 1 ms). Each cell in the TPM gives the probability of a state transitioning to another state. From the displayed 20 samples, the state 00 never transitions to itself in the following sample, thus the transition probability of 00 to 00 is zero. (e) Φ is calculated from a given state and a TPM. Thus, Φ is calculable at every sample.

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 (either 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 2, 3, and 4 channels out of 15 channels, giving 105, 455, and 1365 channel sets respectively per fly (*N* = 13 flies). 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 rapidly. To avoid assuming specific effective connectivity among channels, all channels within a network were considered to be fully connected (i.e., each channel was considered bidirectionally connected to every other channel). Considering full connectivity in the network in this manner forces the search for the MIP to take place over all possible partitioning schemes, instead of a reduced set of partitioning schemes as deduced by ad-hoc assumptions of connectivity. 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”). In this manner, the TPM encodes the causal connections among network elements. 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. In the latest derivation of Φ, this is done by calculating the earth mover’s distance (EMD; {}) between the probability distributions given in the TPM generated by a full network and those given in the TPM generated by the split network. Figure X illustrates this concept.



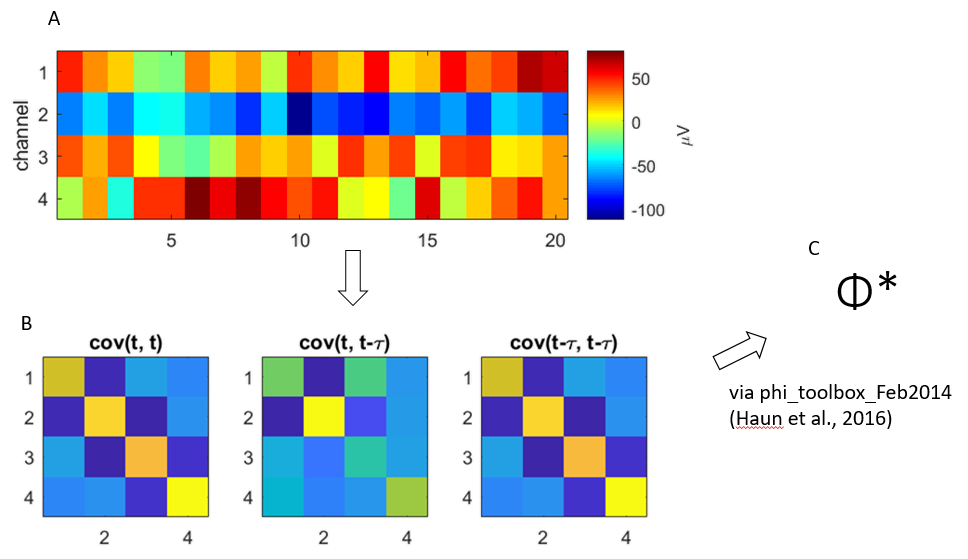
*Figure X.* Φ is calculated by comparing transition probabilities of a split network to a whole network. (a) Discretised time samples as in Figure X. (b) The TPM of the full network, as in Figure X. (c-d) Splitting the network of two channels into two networks of one channel each gives two TPMs, each holding the transition probabilities of their respective part of the network. Multiplying the transition probabilities between parts of the networks gives the joint probability of independent events. This provides the TPM for all parts of the split network, as in (d). In order to calculate Φ, transition probabilities in the full network TPM and the split network TPM are compared by finding the EMD between the TPMs.

While IIT’s exclusion principle advocates for calculation of Φ at the optimal temporal resolution, it is unclear what this resolution is. I thus calculated transition probabilities at three τ levels: 4, 8, and 16 ms. To ensure adequate sampling to obtain accurate transition probabilities, TPMs were built at each of the air and isoflurane conditions using all sample transitions (*N* = 2250 – τ per trial) across all trials.

Collapsing across samples. As each time sample gives a network state, Φ is calculable for every time sample. To obtain an overall Φ value for a trial of 2250 samples, I calculated Φ for every possible network state, then averaged across states, weighting by the number of occurrences of each state within the trial, as in {Albantakis, 2014 #35}. As MIPs are not quantities and thus cannot be averaged, I kept a count of each MIP which occurred in the trial.

### Φ\* Computation

Data processing for computing Φ\* was conducted using MATLAB R2016a in MASSIVE. To calculate Φ\*, I used a toolbox which implemented its calculation in a previous study {<https://github.com/amhaun01/phipattern>, Haun}. The mathematical details for its calculation are provided in {Oizumi, 2016 #46}. As inputs, it takes covariances among continuous signals. Thus, preprocessed LFPs were not discretised as for the calculation of Φ. Unlike PyPhi, the Φ\* toolbox does not search for the MIP, thus it also requires a partitioning scheme as an input. Conceptually, while Φ is given by the distance between two sets of probability distributions, Φ\* is the difference between the information generated by a full network and the information generated by a split network. In the following subsections, I describe how covariances among signals was obtained, and how I obtained MIPs for Φ\*. Network selection followed the same paradigm as Φ computation. The overall pipeline for computing Φ\* is presented visually in Figure X.



*Figure X.* Processing pipeline for calculating Φ\*. (a) LFP samples for a network of four channels. (b) From left to right, covariances across channel signals, without lag, covariances between channel signals and time lagged signals (τ = 4 ms), and covariances between time lagged signals. (c) Φ\* is calculated from covariances. Thus, each trial is associated with a single Φ\* value.

Covariances. Instead of using transition probabilities, Φ\* is calculated using covariances among time series data. Φ\* requires three sets of covariances among all network elements: (a) covariances between signals, (b) covariances between signals and time lagged signals τ, and (c) covariances between time lagged signals. Time lagged signals refer to signals which are offset by some τ. For example, for a signal *tstart* to *tend*, the corresponding time lagged signal is from *tstart* + τ to *tend* + τ. As for TPMs, I calculated covariances at 3 τ levels: 4, 8, and 16 ms. To utilise all sample recordings in a trial, the length of signals (number of samples) was varied to accommodate τ lag. For example, at τ = 4 ms, signals comprised of 2250 – 4 = 2246 samples.

MIP search. The Φ\* toolbox used does not search for the MIP across which to calculate Φ\*. Consequently, I calculated Φ\* for every partitioning scheme. Following the procedure in {Haun}, the partition which produced the minimum normalised Φ\* value {Haun, Balduzi & Tononi 2008} was selected as the MIP, and its unnormalized Φ\* value as the Φ\* value for the network. Thus, while for Φ each trial is associated with multiple MIPs, for Φ\* a trial is associated with only one MIP.

### Data Analysis

Statistical analyses were conducted using MATLAB R2017a.

Air versus isoflurane. To address my first hypothesis that Φ would be reduced under anaesthesia, I employed linear mixed effects (LME) analysis {} to test for effects of condition (air or isoflurane) on Φ values. As I calculated Φ at three network sizes (two, three, and four), and at three τ lags (4, 8, or 16 ms), I also tested for effects of network size and τ lag. Thus, after averaging across trials, 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. Thus, every network, for every fly, was included in this analysis. Fixed effects were tested using likelihood ratio tests between the full model and a null model with the effect of interest removed.

To address my second hypothesis that MIP cuts would be more likely to be feedback cuts under anaesthesia, I looked at the unidirectional MIP cuts in networks of only two channels. This analysis was limited to these networks as previous feedback analysis on the same dataset by {Cohen, #2} was conducted on pairs of channels. Furthermore, in networks of two channels, the MIP contains only one cut (the cut is either to the connection from one channel to the other, or vice versa). Following the same channel grouping scheme as used by {Cohen, #2}, I grouped channels 2-7 as peripheral, and channels 10-15 as central, where the first channel is outermost from the centre of the brain. All other channels were ignored. A feedback connection was defined as a connection from a central channel to a periphery channel. Thus, I considered a MIP cut to be a feedback cut if it severed the connection from a central channel to a peripheral channel. I took the portion of samples within a trial for which the MIP cut was a feedback cut as a measure of feedback influences. After averaging across trials and networks, I used a two-way ANOVA to find effects of condition and lag on feedback influence.

Φ\* versus Φ. To compare Φ\* to Φ, I once again employed LME analysis to test for effects of condition, network size, and τ lag on Φ\* values. Thus, similarly to Φ, I modelled trial-averaged Φ\* as dependent on these effects, and included random intercepts for fly and the interaction between fly and network to account for the nesting of networks within flies. Once again, likelihood ratio tests 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 at each network size and τ lag. Correlations in this manner were calculated across networks. I also assessed MIP equivalence between Φ and Φ\*. To allow for comparability between Φ and Φ\* MIPs, I ignored the directionality of Φ MIP cuts, and excluded trials in which the Φ\* MIP was not a bipartition. Consequently, Φ and Φ\* MIPs will always be identical for networks of two channels (as there is only one way to partition such a network when ignoring directionality), and thus I limited this analysis to networks of three and four channels. I took the portion of samples within a trial whose Φ MIPs matched the Φ\* MIP of the trial as a measure of MIP equality. After averaging matched portions across trials and networks, I conducted one-sample *t*-tests comparing match portions to the expected portion of matches from chance (the conditional probability of a Φ MIP matching a Φ\* MIP is 1/3 and 1/7 for three and four channels respectively, given there are three ways to partition a set of 3 elements and 7 ways to partition a set of four elements when ignoring directionality) at each network size, condition, and τ lag.

## Results

### Φ is Reduced Under Anaesthesia

I conducted LME analysis to compare Φ values between the air and isoflurane conditions. EMD as a distance metric is always zero or greater, thus Φ values were always positive. As Φ values were positively skewed for all flies, I log-transformed trial averaged Φ values before fitting the model to address heteroscedasticity. Figure X displays the change in non-transformed Φ values from the air condition to the isoflurane condition at one τ lag. Across network sizes and τ lags, Φ was significantly reduced in the isoflurane condition, χ2(1) = 2.16 × 104, *p* < .001.

Having calculated Φ at three network sizes and three τ lags, I also tested for effects of these on Φ values. Φ increased significantly with network size (χ2(1) = 1.62 × 104, *p* < .001), and decreased significantly with longer lags (χ2(1) = 8.99 × 103, *p* < .001). To compare decreases in Φ at the channel level at each combination of network size and τ lag, I used post-hoc Wilcoxon signed-rank tests on untransformed values at each network (with FDR corrected *p* < .05). The portion of networks which experienced a significant decrease Φ in this manner increased with the network size, but was relatively stable across τ lags (Figure Xd).

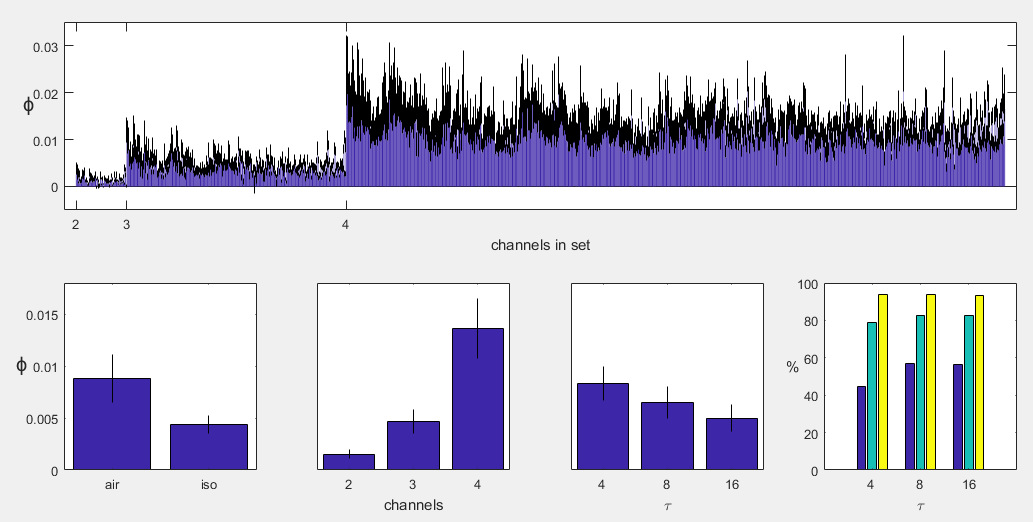


FIGURE X *Figure X.* Average change in Φ values (air – iso) across flies (*N* = 13) (a) Change in Φ for all channel sets at τ = 4 ms. Error bars represent standard error across flies. (b-d) Φ values after averaging across all channel sets within flies, and (b) across set sizes and lags (effect of condition), (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.

### Φ MIPs Do Not Reflect Loss of Feedback

To assess if feedback influences as captured by unidirectional MIP cuts were changed between the air and isoflurane conditions, I conducted a two-way ANOVA with a main effect of condition, and to account for having calculated Φ at three lags, τ lag. The portions of feedback cuts, after averaging across trials and networks, did not significantly change for either condition (*F*(1, 72) = 0.05, *p* = 0.82) or lag (*F*(2, 72) = 1.64, *p* = 0.20). Considering that, for two channels, there are only two possible MIP cuts, the portions of feedback cuts in both the air and isoflurane conditions were also not significantly different from the chance level of 50%. (see Figure X, one sample *t*-tests, *p* > .05 for both conditions at all τ lags).

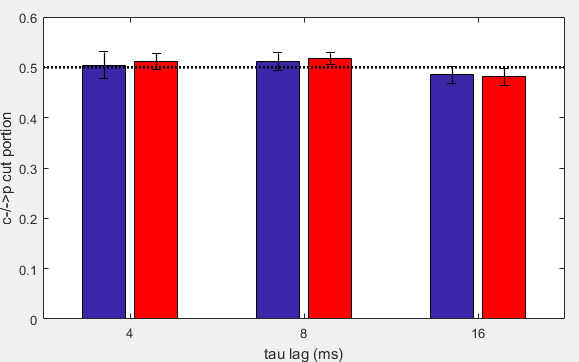


Figure X *Figure X.* Proportion of feedback cuts in the MIP for channel sets of size two. Blue bars indicate the proportion in the air condition, and red bars indicate the proportion in the isoflurane condition. The dotted line represents the proportion of feedback cuts expected by chance (50%).

### Φ\* is Reduced Under Anaesthesia

To compare Φ\* to Φ, I repeated the same statistical analyses for Φ\* as for Φ. Thus, I conducted LME analysis to test for effects of condition, network size, and τ lag. Once again, all Φ\* values were positive, and the distribution of Φ\* within flies was positively skewed. Thus, trial averaged Φ\* values were log transformed before fitting the model to address heteroscedasticity. Figure X displays the change in non-transformed Φ values from the air condition to the isoflurane condition at one τ lag. As for Φ, Φ\* was significantly reduced in the isoflurane condition (χ2(1) = 1.67 × 104, *p* < .001), increased with the number of channels considered (χ2(1) = 1.19 × 104, *p* < .001), and decreased with longer lags (χ2(1) = 481.64, *p* < .001).

I once again used post-hoc Wilcoxon signed-rank tests on untransformed values at each network (with FDR corrected *p* < .05) to compare the portions of networks which experienced a significant decrease in Φ under the isoflurane condition across network sizes and τ lags (Figure Xd). Overall, a smaller portion of networks experienced a decrease in Φ\* than in Φ. As for Φ, the portion of networks which had reduced Φ\* during isoflurane increased with network size. Unlike Φ, the portion of networks with significantly reduced Φ\* was noticeably larger for τ = 4 ms.

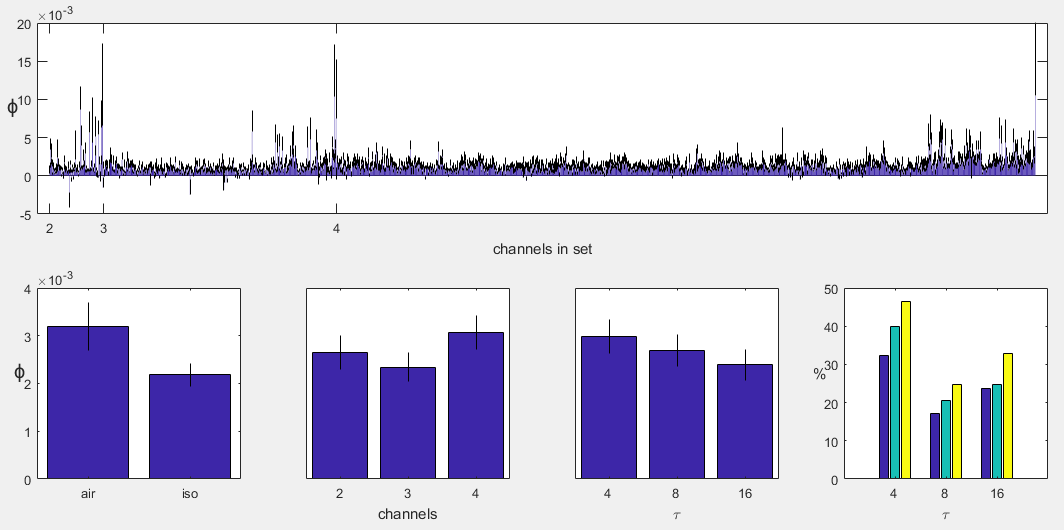


FIGURE X *Figure X.* Average change in Φ\* values (air – iso) across flies (*N* = 13) (a) Change in Φ\* for all channel sets at τ = 4 ms. (b-d) Φ\* values after averaging across all channel sets within flies, and (b) across set sizes and lags (effect of condition), (c) conditions and lags (effect of 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.

### Φ\* is Correlated With Φ

To assess the relationship between Φ\* and Φ values, I calculated correlation coefficients across networks within flies. To account for the three network sizes and three τ lags, I calculated correlations at each network size and τ lag. Figure X displays the relationships between Φ\* and Φ values for one fly at one τ lag. Overall, correlations were generally weak to moderate. Correlations generally were stronger for longer τ lags, for all network sizes.

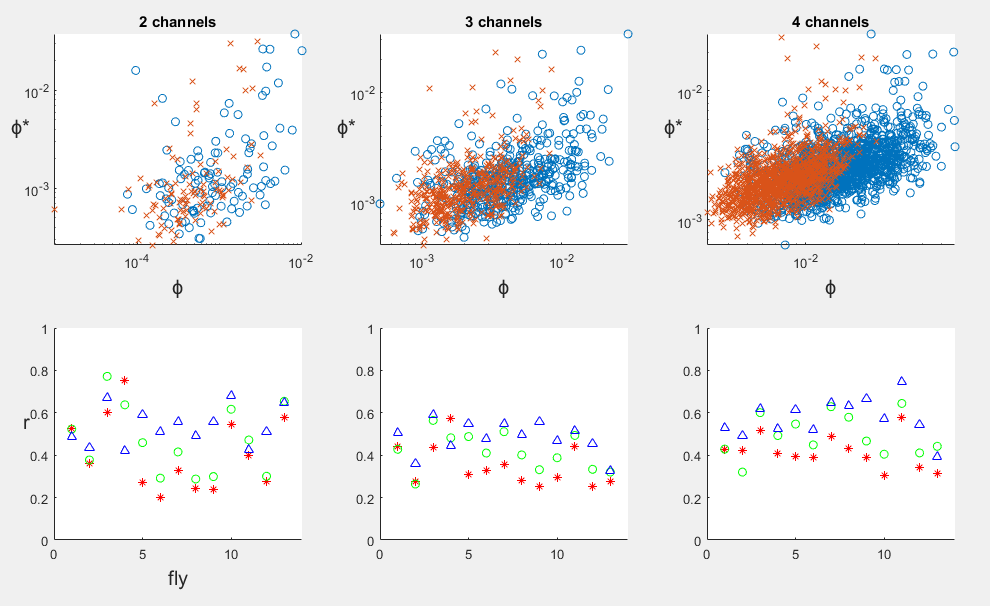
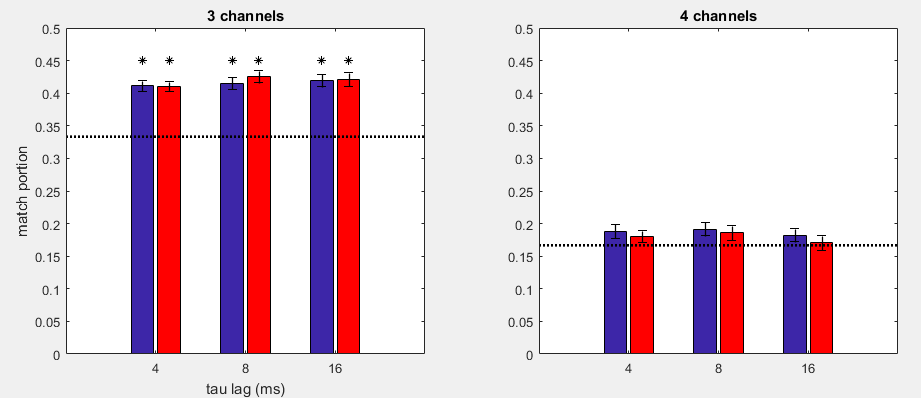


FIGURE X *Figure X.* Relationship between Φ and Φ\*. (a-c) Relationship between Φ and Φ\* for one fly at 4 ms, at each network size. Each point is the trial-averaged Φ and Φ\* value for a channel set. Orange x’s are during isoflurane, and blue circles are during the air condition. (d-f) Φ\* correlation 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 Do Not Match Φ MIPs

To compare MIP equality between Φ\* and Φ, I conducted one-sample *t*-tests at each condition and τ lag, comparing trial-averaged match portions to the match portion expected by chance at each network size (1/3 and 1/7 for three channels and four channels respectively). The portion of matching MIPs for networks of three channels was significantly greater than chance for both conditions and all τ lags (*p* < .05 after Bonferroni correction; see Figure X). However, the portion of matches was no different from chance for networks of four channels, at any condition or τ lag.

To test for differences in trial-averaged match portions due to either condition or τ lag, I also conducted two-way ANOVAs at each network size. These revealed no significant effects of either condition (*F*(1, 72) = 0.25, *p* = 0.62, and *F*(1, 72) = 1.02, *p* = 0.32, for three and four channels respectively) or lag (*F*(2, 72) = 0.74, *p* = 0.48, and *F*(2, 72) = 0.73, *p* = 0.49 for three and four channels respectively) on the portion of MIP matches per trial for both networks of three channels and four channels.



*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**

## Discussion

In this project, I aimed to investigate integrated information Φ as a potential measure of conscious level, and to compare it to a derivative version, Φ\*. To achieve my aims, I calculated both Φ and Φ\* across sets of recordings taken from the fly brain during wakefulness, and during administration of isoflurane anaesthesia. The results indicate that Φ changes with conscious level. To my knowledge, this research is the first to calculate the latest derivation of Φ in a biological system. Consequently, it is also the first to compare it across varied conscious level, and the first to compare it with a derivative measure, Φ\*, in a biological system.

To achieve my aims, I calculated both Φ and Φ\* across sets of recordings taken from the fly brain, during and without the administration of isoflurane anaesthesia. As predicted, Φ was reduced during anaesthesia for the majority of channel groupings. This reduction occurred regardless of the number of channels being assessed, and regardless of the time resolution used to calculated Φ. This reduction was unexpectedly not linked with an increased likelihood of MIPs consisting of a feedback cut, however. Finally, though the effects of anaesthesia on Φ\* was not as clear as for Φ, the derivative measure was moderately correlated with the original measure. To my knowledge, this research is the first to calculate the latest derivation of Φ in a biological system, and thus also the first to compare it across varied conscious level and the first to compare it with a derivative measure, Φ\*, in a biological system.

### Φ decreases during anaesthesia

IIT predicts that its measure, Φ, 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. The results support this hypothesis: Φ calculated from recordings taken during anaesthesia were overall reduced compared to Φ calculated from recordings taken during wakefulness. A larger portion of networks experienced a significant decrease in in larger networks. Though not all networks experienced a decrease in Φ under anaesthesia, Φ for all networks of four channels decreased during anaesthesia. This result adds to the known studies which have also compared Φ between levels of consciousness in a biological system {Lee, 2009 #44}{Chang, 2012 #88}, both of which used a previous derivation of integrated information, and reported reduced Φ during reduced conscious levels.

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. <Interaction between tau and spatial scale of network (e.g. channels are far away might need longer tau, channels close together need shorter tau)>

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.

These results support Φ in its ability to determine level of consciousness in vegetative state etc patients… PCI however was not derived to be able to assess the contents of consciousness. One of the largest advantages of Φ is its ability to also assess contents of consciousness (however, not covered in this thesis).

The results here have demonstrated Φ’s applicability to a non-vertebrate model, specifically the fly. In addition to anaesthetics influence the fly brain similarly to in mammals, Φ behaving as predicted in the fly model further supports the usefulness of the fly model in studying consciousness. It would be interesting to further extend Φ to other animals. Nematodes and roundworms for example exhibit only simple behaviours (e.g. …), and thus Φ may be low in these invertebrates. Φ can be interpreted as representing the “amount” of experiences that a being can have – reduced Φ, as in anaesthesia, indicates the loss of capacity for the system to represent a large set of experiences. In this manner, extending Φ to different animals may allow for the comparison of the number of experiences each may have. However, given that in principle a greater quantity of neurons can hold more information than fewer neurons, it is likely that Φ will be linked to the number of neurons in the animal brain. <However, this also depends on the complex, which is not part of this thesis>

Φ also has utility in development of artificial intelligences (AIs). AIs are currently limited in that they are only able to perform a specific task. In this context, IIT suggests that the purpose of consciousness is to allow an entity with limited resources to adapt to a multiple environments and problems. Already, simulation studies have demonstrated that entities with the same number of elements have increased Φ if they are able to perform more complicated tasks. Thus, future development of AIs may look towards building networks which generate high levels of Φ.

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