Integrated Information Φ in Flies Is Reduced Under Anaesthesia

Angus Leung

Word Count: 9550

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

*Bachelor of Science (Honours)*

*School of Psychological Sciences, Monash University*

*October 2017*Table of Contents

[Abstract 5](#_Toc495675770)

[Statement of Contribution 6](#_Toc495675771)

[CHAPTER 1: INTRODUCTION 7](#_Toc495675772)

[The Search for the Neural Substrate of Consciousness 7](#_Toc495675773)

[The Integrated Information Theory of Consciousness 9](#_Toc495675774)

[Integration, Information, and Conscious Level 14](#_Toc495675775)

[Practicality of Φ as a Measure of Conscious Level 15](#_Toc495675776)

[Testing the IIT in the Fly Model 17](#_Toc495675777)

[Aims and Hypotheses 18](#_Toc495675778)

[CHAPTER 2: EXPERIMENTAL METHODS AND RESULTS 20](#_Toc495675779)

[Method 20](#_Toc495675780)

[Experimental Procedure 20](#_Toc495675781)

[Animal preparation. 20](#_Toc495675782)

[Electrode probe insertion. 20](#_Toc495675783)

[Isoflurane delivery. 21](#_Toc495675784)

[Experimental protocol. 21](#_Toc495675785)

[Local field potential preprocessing. 21](#_Toc495675786)

[Φ Computation 22](#_Toc495675787)

[Discretisation. 23](#_Toc495675788)

[Network Selection. 23](#_Toc495675789)

[Transition Probability Matrix Construction. 23](#_Toc495675790)

[Collapsing across samples. 25](#_Toc495675791)

[Φ\* Computation 25](#_Toc495675792)

[Covariances. 26](#_Toc495675793)

[MIP search. 27](#_Toc495675794)

[Data Analysis 27](#_Toc495675795)

[Air versus isoflurane. 27](#_Toc495675796)

[Φ\* versus Φ. 28](#_Toc495675797)

[Results 30](#_Toc495675798)

[Φ During Wakefulness and Anaesthesia 30](#_Toc495675799)

[Φ MIP Feedback 31](#_Toc495675800)

[Φ\* During Wakefulness and Anaesthesia 32](#_Toc495675801)

[Φ\* Versus Φ 33](#_Toc495675802)

[Φ\* MIPs Versus Φ MIPs 34](#_Toc495675803)

[CHAPTER 3: DISCUSSION AND CONCLUSION 36](#_Toc495675804)

[Discussion 36](#_Toc495675805)

[Φ Decreases During Anaesthesia 36](#_Toc495675806)

[Φ Increases with Network Size. 37](#_Toc495675807)

[Φ Decreases at Longer Timescales. 37](#_Toc495675808)

[Φ MIPs Do Not Indicate Reduced Feedback 38](#_Toc495675809)

[Φ\* Versus Φ 39](#_Toc495675810)

[Limitations and Future Directions 41](#_Toc495675811)

[Conclusion 43](#_Toc495675812)

[CHAPTER 4: REFERENCES 44](#_Toc495675813)

# 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 two, three, and four channels, at varying timescales. Both Φ and Φ\* were significantly reduced during anaesthesia, increased with number of channels, and decreased with larger timescales. Overall, the two quantities were moderately correlated. These results support IIT’s prediction of reduced Φ with reduced conscious level, 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.

# 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: 13/10/2017

# **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), high-order thoughts (Edelman, 2003), or reportable access to one’s own experience (Block, 1995). 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, Thorp, Graham, & Aitkenhead, 1991; Sebel et al., 2004). Furthermore, physiologic responses to stress during consciousness under anaesthesia, such as increased heart rate, are often masked by accompanying drugs (Rani & Harsoor, 2012). 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) or physiological signs (Rani & Harsoor, 2012). 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, Massimini, Boly, & Tononi, 2016). 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 & Singer, 2001), 40 Hz oscillations in the cerebral cortex (Llinás & Ribary, 1993), and feedback interactions (Lamme, 2010) have been proposed as parts of level NCC, and activations of cortical areas such as the fusiform face area (FFA) during face perception (Pierce, Müller, Ambrose, Allen, & Courchesne, 2001) have been proposed as parts of content NCC. This reductionist approach to understanding consciousness however has significant drawbacks.

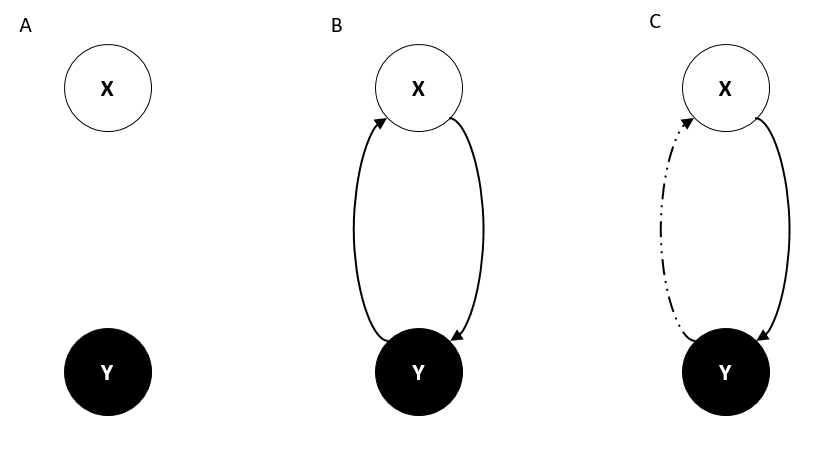
The first drawback is that observed NCC cannot be taken as reliable indicators of consciousness. For example, synchronous activity and feedback interactions both occur in the cerebellum (Person & Raman, 2012; Witter, Rudolph, Pressler, Lahlaf, & Regehr, 2016), which likely does not contribute to consciousness (Yu, Jiang, Sun, & Zhang, 2015), while the FFA is activated during perception of non-face stimuli (Gauthier, Skudlarski, Gore, & Anderson, 2000; Gauthier, Tarr, Anderson, Skudlarski, & Gore, 1999). 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, Hernández-Pérez, & Concha, 2016). 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, 1995), 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 (B. J. Baars, 1997; Bernard J. Baars, 2002) and cross-order integration theory (Kriegel, 2007). 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 (Sergent & Dehaene, 2004) 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, Albantakis, & Tononi, 2014) 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 be 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 et al., 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.

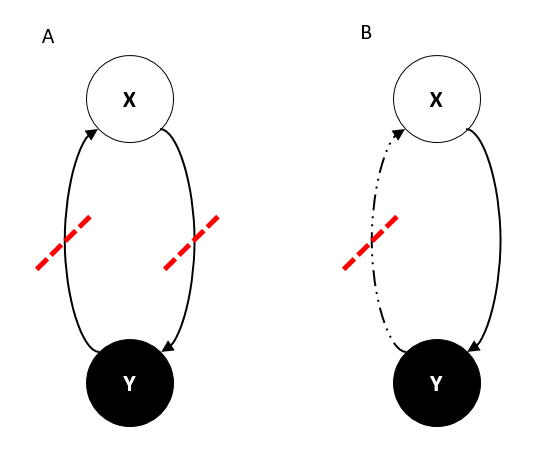
IIT proposes that a conscious network specifies its own causes and effects. That is, in a conscious network, the network state (i.e. the combination of the states of all elements in the network, for example the state of all neurons in the brain) should be informative as to the possible previous network 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 1). For details, its full mathematical derivation, which is unnecessarily complicated for this thesis, is provided in (Oizumi et al., 2014).



*Figure 1*. Assessment of information 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 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* through a causal connection (black arrow), then X’s state at *t* + 1 must be 0 in this example. Similarly, if Y assumes X’s state at *t* through a causal connection, then Y’s state at *t* + 1 must be 1. Thus, at *t* + 1, the network state must be 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.

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 (Figure 1a), 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* + 1 takes the state the other neuron took at time *t* (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 (Figure 1b). 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 (Figure 1c). 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 (Figure 2a), 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 last example however (Figure 2c), 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.



*Figure 2.* Assessment of Φ in networks of two neurons. Networks (a) and (b) are the same as (b) and (c) in Figure 1, respectively. Dotted red lines indicate potential MIP cuts. (b) 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) 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). In this case, the network is only unidirectionally connected (from X to Y).

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 either through only feedforward or only 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 (Ross, Thompson, & Yenkosky, 1997), 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 et al., 2010). The same localisation and stereotyping of responses is observed also dreamless non-rapid eye movement sleep, when compared to rapid eye movement sleep (Massimini et al., 2010), 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 in 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, Mashour, Kim, Noh, & Choi, 2009), and during sleep (Chang et al., 2012). These Φ studies however are based on an older derivation of Φ (Tononi, 2008), 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, Amari, Yanagawa, Fujii, & Tsuchiya, 2016) 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 et al., 2016), 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; Bell, 1934; Toker & Sommer, 2017). 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 & Sommer, 2017).

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 & Seth, 2011) and Φ\* (Oizumi et al., 2016). 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 et al., 2016). 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. Furthermore, when assessing integration, they ignore directionality of causal connections – consequently, instead of assessing how irreducible a network is to subnetworks connected only through unidirectional interactions, they assess how irreducible a network is to independent subnetworks which are not connected at all. 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 & Sommer, 2016). 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 & Sommer, 2017). 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, the 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 test 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 et al., 2012; Herculano-Houzel, Mota, & Lent, 2006). Because of this, mapping of the entire fly brain at the neuronal level is estimated to reach completion within the next 10 years (Alivisatos et al., 2012). In the same timeframe for mammals, one might achieve mapping of only the mouse retina or hippocampus (Alivisatos et al., 2012). 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 invertebrates such as the nematode and roundworm exhibit simple behaviours dependent only on their immediate sensory environment (Barron & Klein, 2016), the fly exhibits selective attention (Sareen, Wolf, & Heisenberg, 2011; van Swinderen, 2005), and spatial memory (Seelig & Jayaraman, 2015). Though these behaviours alone are not sufficient to determine the presence of consciousness, they are generally used to infer consciousness in other animals (Mather, 2008). In addition to these behaviours, flies exhibit alterations in activity, suggestive of altered conscious level, such as torpidness similar to a sleep state in mammals (Hendricks et al., 2000; Shaw, Cirelli, Greenspan, & Tononi, 2000). Just as in mammals, such periods of sleep have distinct states (van Alphen, Yap, Kirszenblat, Kottler, & van Swinderen, 2013).

In addition to complex behaviour and natural states of altered conscious level, administration of anaesthesia seems to affect the fly in a similar as to mammals. Anaesthetics reduce behavioural responses in flies (Allada & Nash, 1993) at similar concentrations required for mammals (van Swinderen, 2006). Furthermore, there is evidence to suggest that the neural mechanisms through which this is achieved is common across animals. For example, loss of communication across cortical regions due to anaesthesia is observed in both mammals (Alkire, Hudetz, & Tononi, 2008) and flies (Cohen, Zalucki, van Swinderen, & Tsuchiya, 2016). Anaesthesia also impairs feedback influences from frontal brain regions to posterior regions (Boly et al., 2012; Ku, Lee, Noh, Jun, & Mashour, 2011), and a similar reduction in feedback has recently been described in the fly (Cohen, van Swinderen, & Tsuchiya, 2017).

Thus, given that flies have significantly fewer neurons than mammals while still displaying a wide range of behaviours, they present an ideal model in which to test tenets of IIT, specifically Φ as a measure of conscious level.

## 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. Additionally, though the approximation of Φ, Φ\*, has previously been calculated across biological recordings (Toker & Sommer, 2017), 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 et al., 2017), 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 also that Φ\* and Φ would be positively correlated. Furthermore, as both Φ and Φ\* assess how a network constrains its past, I hypothesised that the likelihood of Φ\* MIPs matching with Φ MIPs would be greater than chance.

# **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 et al., 2017), 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, Zhou, Stratton, Liu, & van Swinderen, 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 (Cohen et al., 2016) 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 et al., 2013).

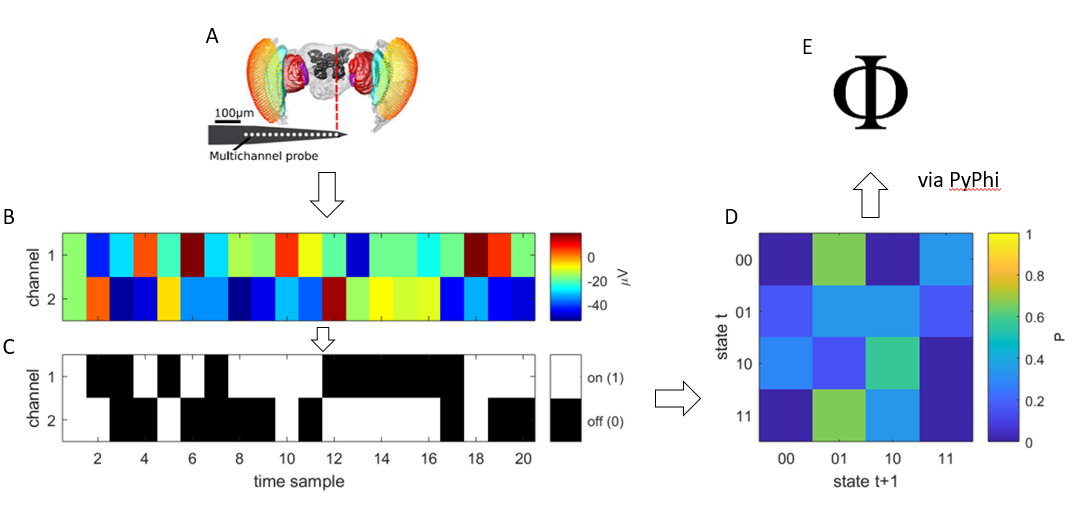
Isoflurane delivery. Isoflurane was delivered from an evaporator onto the fly through a connected rubber hose. The isoflurane was delivered at a constant flow of 2 l/min and continuously vacuumed from the opposite side of the fly. Actual concentration near the fly body was either 0 vol% (air condition) or 0.6 vol% (isoflurane condition) as estimated following a gas chromatography procedure described by van Alphen et al. (2013) for measuring isoflurane concentration. Flies in the air condition responded to air puffs by moving their legs and abdomen, but were rendered inert under the isoflurane condition (Cohen et al., 2016). 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 & 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 used directly 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 et al., 2014). 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 3.

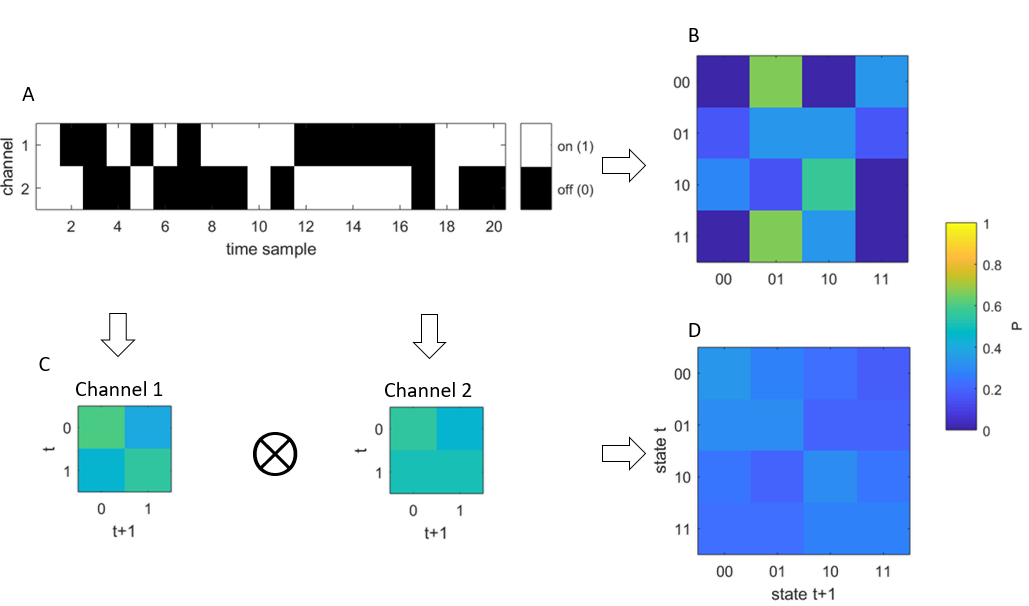


*Figure 3.* My processing pipeline for calculating Φ. (a-b) A linear multi-electrode probe records LFPs from the fly brain. Displayed in (b) are 20 samples for a network of two channels, after pre-processing. (c) Samples in each channel are discretised into one of two values. The state of the network 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; Rubner, Tomasi, & Guibas, 2000) 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 4 illustrates this concept.



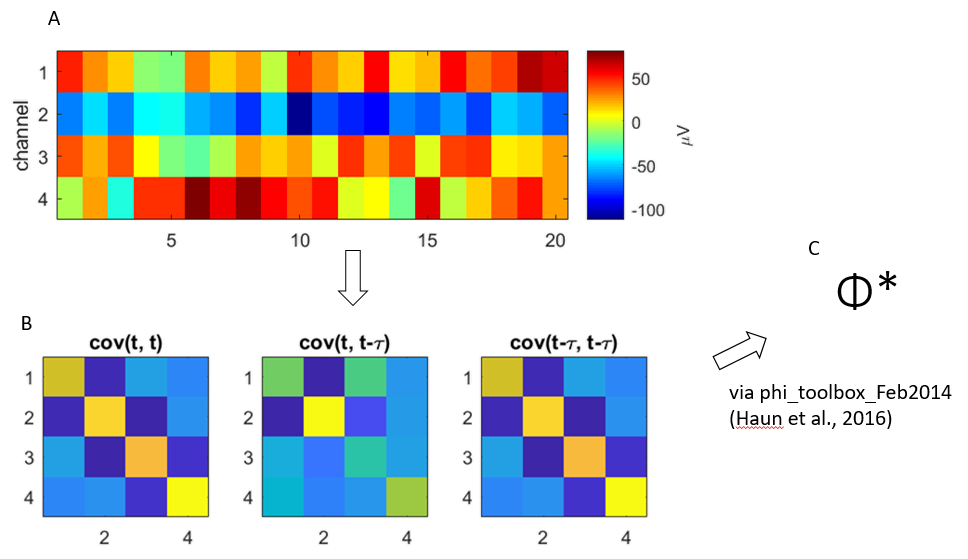
*Figure 4.* Φ is calculated by comparing transition probabilities of a split network to a whole network. (a) Discretised time samples as in Figure 3. (b) The TPM of the full network, as in Figure 3. (c-d) Splitting the network of two channels into two networks of one channel each gives two TPMs, each 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, Hintze, Koch, Adami, & Tononi, 2014). 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 et al., 2016). The mathematical details for its calculation are provided in (Oizumi et al., 2016). As inputs, it takes covariances among 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 5.



*Figure 5.* 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 et al., 2016), the partition which produced the minimum normalised Φ\* value (Equation 1; Balduzzi & 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. Normalised Φ\* is given by:

where *m* is the number of parts in the partition, and is the *k*th part of the partitioned system. is the entropy (an information theoretical measure of baseline uncertainty; (Rényi, 1961) of part .

### 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 (Bates, Mächler, Bolker, & Walker, 2015) 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 et al. (2017) was conducted only 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 et al. (2017), 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 analysis of variance (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 for each fly. To find if correlations were affected by network size or τ lag, I conducted a two-way ANOVA to test for these effects on Fisher’s *z* transformed correlation coefficients (Corey, Dunlap, & Burke, 1998).

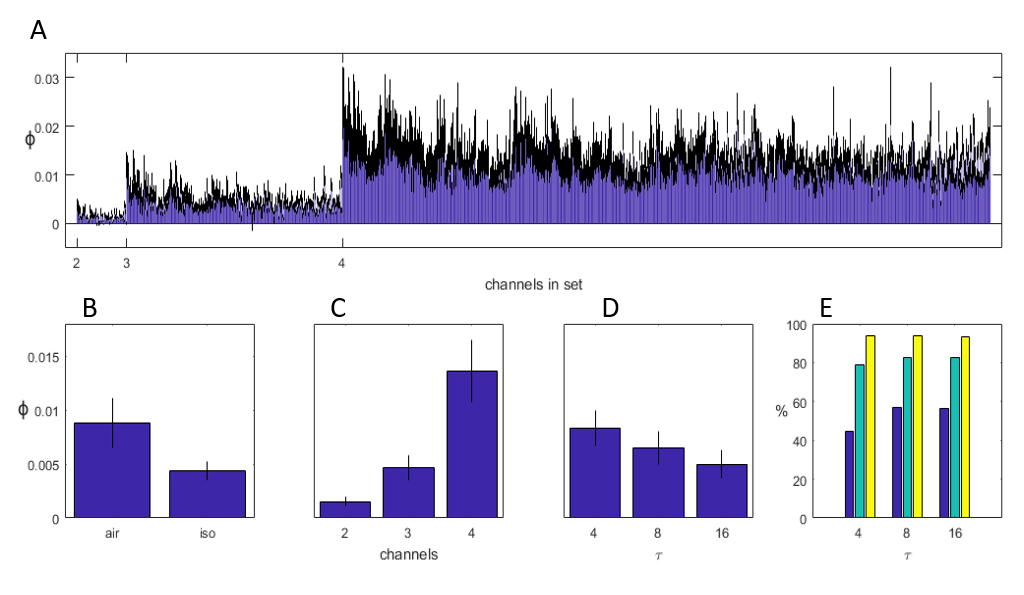
I also assessed MIP equivalence between Φ and Φ\*. To allow for comparability between Φ and Φ\* MIPs, I ignored the directionality of Φ MIP cuts, and excluded trials in which the Φ\* MIP was not a bipartition. Consequently, Φ and Φ\* MIPs will always be identical for networks of two channels (as there 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 During Wakefulness and 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 6 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.

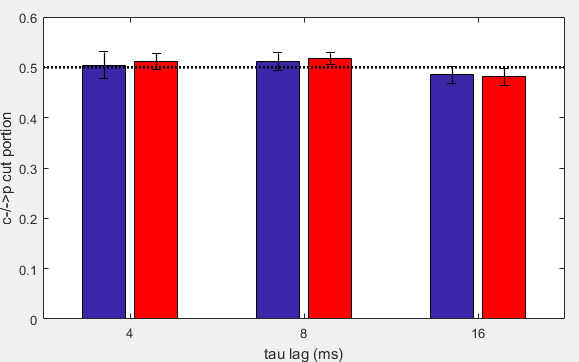
Having calculated Φ at three network sizes and three τ lags, I also tested for effects of these on Φ values. Φ increased significantly with network size (χ2(1) = 1.62 × 104, *p* < .001), and decreased significantly with longer lags (χ2(1) = 8.99 × 103, *p* < .001). To compare decreases in Φ at the channel 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 6e).



*Figure 6.* Φ values in relation to condition, network size, and τ lag. (a) Change in Φ (air – iso) for all networks at τ = 4 ms. Error bars represent standard error across flies (*N* = 13). (b-d) Φ values after averaging across all networks within flies, for (b) effect of condition (averaged across network sizes and lags), (c) effect of network size (averaged across conditions and lags), and (d) effect of lag (averaged across conditions and network sizes). All main effects were significant. (e) Percentage of networks 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 MIP 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 τ lags. 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 7, one sample *t*-tests, *p* > .05 for both conditions at all τ lags).

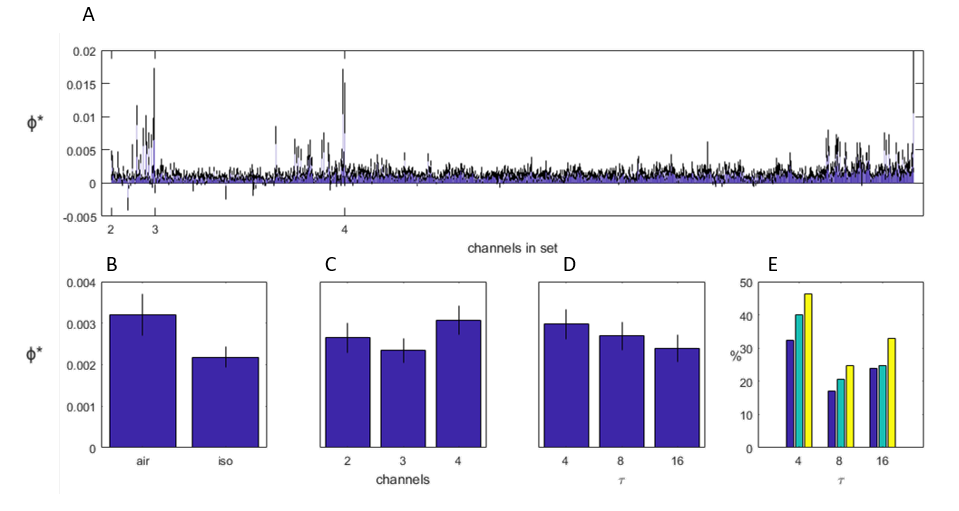


*Figure 7.* Proportion of feedback cuts in the MIP for networks of two channels. Blue bars indicate the proportion in the air condition, and red bars indicate the proportion in the isoflurane condition. The dotted line represents the proportion of feedback cuts expected by chance (50%).

### Φ\* During Wakefulness and 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 8 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).

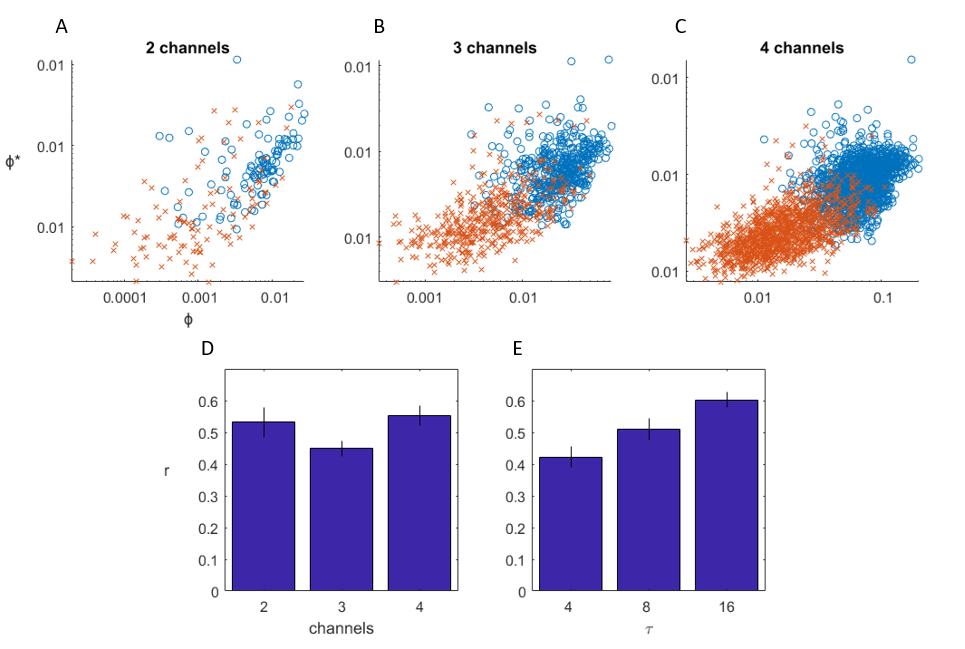
I once again used post-hoc Wilcoxon signed-rank tests on untransformed values at each network (with FDR corrected *p* < .05) to compare the portions of networks which experienced a significant decrease in Φ under the isoflurane condition across network sizes and τ lags (Figure 8e). Overall, a smaller 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 8.* Φ values in relation to condition, network size, and τ lag. (a) Average change in Φ\* (air – iso) for all networks at τ = 4 ms. Error bars represent standard error across flies (*N* = 13). (b-d) Φ\* values after averaging across all networks within flies, for (b) effect of condition (averaged across network sizes and lags), (c) effect of network size (averaged across conditions and lags), and (d) effect of lag (averaged across conditions and network sizes). All main effects were significant. (e) Percentage of networks 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\* Versus 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 9 displays the relationships between Φ\* and Φ values for one fly at one τ lag. Correlations per fly ranged from *r* = .2 to *r* = .77. Correlations generally were stronger for longer τ lags, for all network sizes (Figure 9). I conducted a two-way ANOVA to find effects of network size and τ lag on Fisher *z* transformed correlation coefficients. This revealed significant effects of both size, *F*(2, 116) = 5.23, *p* = .007, and lag, *F*(2, 116) = 13.82, *p* > .001.

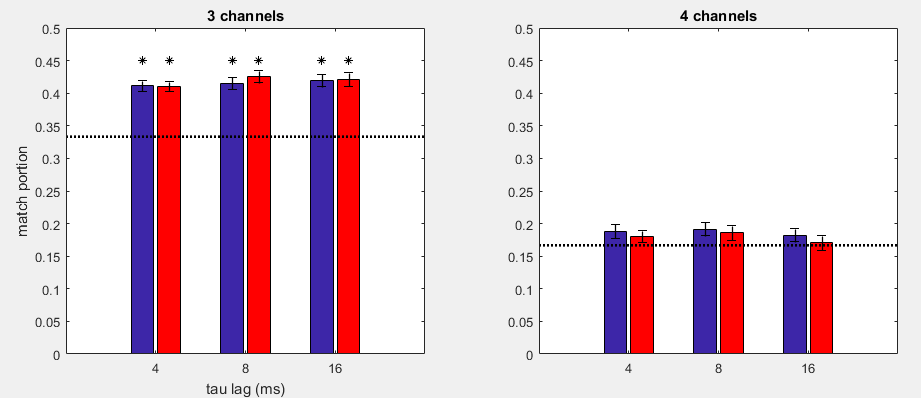


*Figure 9.* Relationship between Φ and Φ\*. (a-c) Relationship between Φ and Φ\* for one fly at 4 ms, at each network size. Each point is the trial-averaged Φ and Φ\* value for a network. Orange x’s are during isoflurane, and blue circles are during the air condition. (d-e) Average Φ\* correlation coefficients to Φ (after Fisher *z* transform and back-transform) for (d) each network size (averaged across τ lags), and (e) each τ lag (averaged across network sizes). Error bars represent standard error across flies (*N* = 13).

### Φ\* MIPs Versus 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 10). 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.



A

B

*Figure 10.* The proportion of Φ MIP cuts in a trial matching that of the Φ\* cut for the trial, averaged across trials, networks, 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 both Φ and Φ\* change 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 levels, and the first to compare it to the derivative measure, Φ\*, in a biological system.

### Φ Decreases During Anaesthesia

IIT proposes Φ as a measure 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. This result adds to the known studies which have also compared Φ between levels of consciousness in a biological system (Chang et al., 2012; Lee et al., 2009), both of which used a previous derivation of integrated information, and reported reduced Φ during reduced conscious levels. Furthermore, while not all networks experienced a decrease in Φ under anaesthesia, increasingly larger portions of networks experienced a significant decrease in larger networks: though only ~50% of networks of two channels experienced a significant reduction in Φ during anaesthesia, more than 90% of networks of four channels experienced this reduction. This indicates that larger biological networks are in a sense more affected (with regards to losing Φ) by anaesthesia than smaller networks. This may be due to larger networks having more causal connections that smaller networks – thus when these connections are interrupted, larger networks experience a larger loss of communication across parts. These results directly support Φ as a measure of conscious level, and supports the general concept of anaesthesia reducing communication among brain regions.

Φ Increases with Network Size. In addition to reducing with reduced conscious level, Φ also increased with number of channels considered. This result may be due simply due to larger quantities of binary elements being able to generate larger quantities of information (increasing the number of binary elements in a network by one doubles the number of states the network can take). Importantly though, increased Φ indicates that the greater amount of information is dependent on interactions among channels which integrate the network. This result is consistent with regards to IIT. Φ as proposed by IIT is purposed to also reflect conscious experience, and in this context its magnitude can be interpreted as a measure of the quantity of possible experiences a conscious system may have. For example, a system consisting of two integrated 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 channels, it is likely that Φ will continue to increase in a biological system as more elements are included, so long as they remain integrated.

Φ Decreases at Longer Timescales. A key question with regards to Φ is the temporal 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 network having 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 finding this scale was not a primary aim of this study, the results indicate that maximal Φ may be attained at a more granular (shorter) timescale. Given that time perception is linked to body size, however (Healy, McNally, Ruxton, Cooper, & Jackson, 2013), this finding of increased Φ at such a short timescale may be limited to small networks or small animals only. It is likely that as more elements and thus interactions are considered, the optimal resolution at which Φ is maximal will increase also.

### Φ MIPs Do Not Indicate Reduced Feedback

General anaesthetics reduce feedback interactions from frontal brain regions to posterior regions (Boly et al., 2012). Recently, Cohen et al. (2017) applied frequency domain Granger causality analysis to the same data set used in the present project, and identified a similar reduction in feedback influences due to anaesthesia, from central channels to peripheral channels. Granger causality is a measure of the extent to which a signal can be used to predict another signal – a loss in feedback influence thus is interpretable as the signal of a central channel losing effectiveness in predicting the signal of a peripheral channel. As the MIP partitions a network by ignoring connections between elements which do not contribute to the information generated by the network, it should thus be the case that MIPs will cut along feedback connections if these connections are uninformative. Thus, it was anticipated that MIP cuts from central channels to peripheral channels would be more likely during anaesthesia. This hypothesis was not supported, however, as the likelihood of having a feedback cut as the MIP cut remained fixed at chance level in both the air and isoflurane conditions.

This result should be interpreted with caution. Reduction in feedback influences by Cohen et al. (2017) in the same dataset was characterised at low frequencies (0-5 Hz), corresponding in principle to a time resolution of roughly 200 ms. I only calculated Φ at resolutions up to 16 ms, which may not be sufficient to characterise the same loss in feedback influences. However, there is no literature demonstrating that frequency domain analyses and time domain analyses can be directly linked together at some timescale, and so it is unclear whether there would be a greater likelihood of feedback cuts if Φ was calculated using longer τ lags.

The TPM presents perhaps an explanation as to the lack of difference in likelihood of feedback cuts between wakefulness and anaesthesia. By holding the transition probabilities of each network state to every other network state, the TPM indirectly holds the causal interactions among network elements. As I calculated separate TPMs for each of wakefulness and isoflurane anaesthesia, it is possible that loss of feedback influences is captured instead in the differences between the TPMs of each condition. Consequently, MIP cuts in the contexts of TPMs which represent different causal connections may not be directly comparable.

### Φ\* Versus Φ

Recent literature involved in investigating integrated information as a potential measure of consciousness have generally utilised a derivative version of Φ, based on a previous iteration of IIT. The derivative version, Φ\* (Oizumi et al., 2016), is appealing as it does not require knowledge of transition probabilities, and thus overcomes the otherwise infeasible observational requirements of Φ in large networks. Furthermore, it is analytically calculable under the assumption of Gaussian variables (Oizumi et al., 2016) and is thus computationally cheaper to calculate than the original Φ. 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 (Tononi, 2008), allowing for easier understanding of integrated information. Despite its popularity however, it has not yet been compared directly to the latest version of Φ. Φ\* meets the theoretical requirements of both the version of IIT upon which it was derived. Furthermore, the latest version of IIT, while updating certain concepts, is conceptually similar to its previous iteration. Thus, 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\*. I also expected above chance likelihood of Φ MIPs matching Φ\* MIPs. The results partially support convergent validity between Φ\* and Φ. As anticipated, Φ\* behaved similarly to Φ – it was reduced under anaesthesia, and overall increased with the network size and decreased with longer τ lags. However, Φ and Φ\* MIPs matched only at chance level.

Though Φ\* behaved similarly to Φ with regards to condition, network size, and timescale, Φ\* values overall were reduced in relation to Φ. This may reflect the theoretical differences in the versions of IIT upon which Φ\* and Φ are built. One of these is the assessment of information by considering only how a network constrains its past states (Tononi, 2008), as opposed to both its past and future states (Oizumi et al., 2014). A consequence of this is that there is inherently more uncertainty in a network when assessing it using Φ – there is uncertainty as to both the possible past states and the possible future states. Thus, assessing both the past and future, as Φ does, allows for more information to be generated, when compared to assessing either only the past or only the future. A second difference lies in the actual calculation of information. Φ\* calculates information using Kullback-Leibler (KL) divergence (Oizumi et al., 2016), which does not account for similarity among network states. Similarity refers to how different two states are – for example, the state 11 might be considered more similar to the state 10 than to the state 00. Meanwhile, Φ calculates information using the EMD (Oizumi et al., 2014). The EMD is able to describe constraints among dissimilar state as stronger, which leads to the calculation of a larger quantity of information. Consequently, the use of the EMD may also contribute to the greater magnitude of Φ values over Φ\*.

Positive correlations between Φ\* and Φ give promise to Φ\* as a practical version of Φ. However, fewer than half of all networks experienced a significant decrease in Φ\* during anaesthesia, compared to over 90% for Φ (Figures 6e and 8e). Thus, while Φ\* may indicate Φ in a network, it may not be ideal for determining changes in conscious level. Interestingly, though both Φ and Φ\* decreased with longer time lags, the correlation between Φ and Φ\* was on average strongest at the longest timescale of 16 ms. It may be the case the Φ\* behaves more similarly to Φ at longer timescales. Despite the consistent correlations between Φ and Φ\*, the likelihood of MIPs matching, however, remained close to chance. Even though correlations increased at longer timescales, MIP match likelihood remained constant at all timescales. As with reduced magnitude of Φ\* values in relation to Φ values, the lack of MIP matching may also be a reflection of the theoretical differences between Φ\* and Φ. Specifically, Φ’s additional assessment of how a network constrains its future may result in different MIPs from MIPs which result solely from assessment of how a network constrains its past.

### Limitations and Future Directions

Given that the latest derivation of Φ has not been applied to biological recordings before now, some methodological decisions made in the calculation of Φ were made without the direct support of past literature. This leads to several overarching limitations for this study. Here I review these limitations consequently suggest future avenues with which to take the investigation of Φ.

The first methodological decision involves the discretisation of continuous samples to accommodate the calculation of Φ. As Φ requires discrete variables, the data used to calculate Φ was discretised into binary variables, based on the median sample value. Whether this is the ideal method with which to make data compatible with Φ however is unclear, as one may discretise recordings in different ways. For example, King et al. (2013) introduces the symbolic transform, which converts the pattern of a contiguous set of samples to a discrete value. One advantage of the simple discretisation method used in this project is its extendibility. Discretisation based on thresholds can in principle be extended continuously – as more thresholds are included, the set of discrete values which a sample can take increases in size. In the case where there are infinite thresholds, discretised samples would thus be no different from continuous samples. Ultimately though, how to best discretise continuous recordings for calculation of Φ is not yet investigated, and whether the behaviour of Φ changes depending discretisation method is unclear. To address this, comparative studies investigating how Φ changes with discretisation methods are recommended. As there are any number of arbitrary ways to discretise recordings, however, perhaps a more promising approach to address this might involve deriving a measure which is applicable to continuous variables, like Φ\*, for the latest iteration of IIT.

The second methodological decision involves calculating Φ\* across continuous, rather than discrete variables. While the appeal of Φ\* is in its applicability to continuous signals, it is not limited to these. As it is calculated from covariances, it can also be applied to discrete variables. Consequently, it is possible that reported dissimilarities between Φ\* and Φ may be due to the differences in the variables across which Φ\* and Φ were calculated. To address this, a future project might aim to compare Φ\* to Φ using only discretised recordings for both versions.

A final methodological decision involves calculating Φ for networks of only up to four channels. This limit was chosen to take into account the costly search for the MIP. While a good first step for IIT, a network of four elements is however very simple, compared to a network of, for example, 15 channels. Thus, it is unclear whether the results described in this study extend to networks of five or more channels. To extend Φ to larger networks, however, it is necessary to be able to quickly search for the MIP. In this study, how channels were causally connected was unclear, as LFPs collate across neural activity. Thus, the search for the MIP was conducted across all possible bipartitions. However, as knowledge of such causal connections reduces the search space for the MIP (Toker & Sommer, 2017), future studies in the fly model may aim to use smaller scale recordings, such as single unit activity, and take into consideration the fly connectome to allow for faster calculation of Φ, and for calculation of Φ in larger networks.

## Conclusion

Results from this project demonstrate that the theoretically derived measure of consciousness, Φ, is reduced in the fly brain during anaesthesia. However, there was no evidence to suggest that this reduction is due to reduced feedback influences in the brain. This project additionally demonstrates that a more practical version of Φ, Φ\*, is also reduced in the fly brain during anaesthesia, and is furthermore correlated with Φ. This is despite Φ\* MIPs and Φ MIPs not matching at a greater than chance level. These findings provide basic evidence which supports IIT, and demonstrates the applicability of both Φ and Φ\* to the fly model. This opens up the potential to further assess Φ, and even test other tenets of IIT, using the fly model.

## CHAPTER 4: REFERENCES

Aitken, A. (1933). A problem in combinations. *Edinburgh Mathematical Notes, 28*, xviii-xxiii.

Albantakis, L., Hintze, A., Koch, C., Adami, C., & Tononi, G. (2014). Evolution of Integrated Causal Structures in Animats Exposed to Environments of Increasing Complexity. *PLOS Computational Biology, 10*(12), e1003966. doi:10.1371/journal.pcbi.1003966

Alivisatos, A. P., Chun, M., Church, George M., Greenspan, Ralph J., Roukes, Michael L., & Yuste, R. (2012). The Brain Activity Map Project and the Challenge of Functional Connectomics. *Neuron, 74*(6), 970-974. doi:10.1016/j.neuron.2012.06.006

Alkire, M. T., Hudetz, A. G., & Tononi, G. (2008). Consciousness and Anesthesia. *Science (New York, N.Y.), 322*(5903), 876-880. doi:10.1126/science.1149213

Allada, R., & Nash, H. A. (1993). Drosophila melanogaster as a Model for Study of General Anesthesia: The Quantitative Response to Clinical Anesthetics and Alkanes. *Anesthesia & Analgesia, 77*(1), 19-26.

Baars, B. J. (1997). In the theatre of consciousness. Global Workspace Theory, a rigorous scientific theory of consciousness. *Journal of Consciousness Studies, 4*(4), 292-309.

Baars, B. J. (2002). The conscious access hypothesis: origins and recent evidence. *Trends in Cognitive Sciences, 6*(1), 47-52. doi:10.1016/S1364-6613(00)01819-2

Balduzzi, D., & Tononi, G. (2008). Integrated Information in Discrete Dynamical Systems: Motivation and Theoretical Framework. *PLOS Computational Biology, 4*(6), e1000091. doi:10.1371/journal.pcbi.1000091

Barrett, A. B., & Seth, A. K. (2011). Practical Measures of Integrated Information for Time-Series Data. *PLOS Computational Biology, 7*(1), e1001052. doi:10.1371/journal.pcbi.1001052

Barron, A. B., & Klein, C. (2016). What insects can tell us about the origins of consciousness. *Proceedings of the National Academy of Sciences, 113*(18), 4900-4908.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *2015, 67*(1), 48. doi:10.18637/jss.v067.i01

Bell, E. T. (1934). Exponential Polynomials. *Annals of Mathematics, 35*(2), 258-277. doi:10.2307/1968431

Block, N. (1995). On a confusion about a function of consciousness. *Behavioral and Brain Sciences, 18*(2), 227-247.

Boly, M., Moran, R., Murphy, M., Boveroux, P., Bruno, M.-A., Noirhomme, Q., . . . Friston, K. (2012). Connectivity changes underlying spectral EEG changes during propofol-induced loss of consciousness. *The Journal of Neuroscience, 32*(20), 7082-7090. doi:10.1523/JNEUROSCI.3769-11.2012

Chalmers, D. J. (1995). Facing up to the problem of consciousness. *Journal of Consciousness Studies, 2*(3), 200-219.

Chang, J.-Y., Pigorini, A., Massimini, M., Tononi, G., Nobili, L., & Van Veen, B. D. (2012). Multivariate autoregressive models with exogenous inputs for intracerebral responses to direct electrical stimulation of the human brain. *Frontiers in Human Neuroscience, 6*, 317. doi:10.3389/fnhum.2012.00317

Cohen, D., van Swinderen, B., & Tsuchiya, N. (2017). Isoflurane reduces feedback in the fruit fly brain. *bioRxiv*, 161976.

Cohen, D., Zalucki, O. H., van Swinderen, B., & Tsuchiya, N. (2016). Local Versus Global Effects of Isoflurane Anesthesia on Visual Processing in the Fly Brain. *eNeuro, 3*(4), ENEURO.0116-0116.2016. doi:10.1523/ENEURO.0116-16.2016

Corey, D. M., Dunlap, W. P., & Burke, M. J. (1998). Averaging Correlations: Expected Values and Bias in Combined Pearson rs and Fisher's z Transformations. *The Journal of General Psychology, 125*(3), 245-261. doi:10.1080/00221309809595548

Cuaya, L. V., Hernández-Pérez, R., & Concha, L. (2016). Our Faces in the Dog's Brain: Functional Imaging Reveals Temporal Cortex Activation during Perception of Human Faces. *PLOS ONE, 11*(3), e0149431. doi:10.1371/journal.pone.0149431

Edelman, G. M. (2003). Naturalizing consciousness: A theoretical framework. *Proceedings of the National Academy of Sciences, 100*(9), 5520-5524. doi:10.1073/pnas.0931349100

Engel, A. K., & Singer, W. (2001). Temporal binding and the neural correlates of sensory awareness. *Trends in Cognitive Sciences, 5*(1), 16-25. doi:http://dx.doi.org/10.1016/S1364-6613(00)01568-0

Ferrarelli, F., Massimini, M., Sarasso, S., Casali, A., Riedner, B. A., Angelini, G., . . . Pearce, R. A. (2010). Breakdown in cortical effective connectivity during midazolam-induced loss of consciousness. *Proceedings of the National Academy of Sciences of the United States of America, 107*(6), 2681-2686. doi:10.1073/pnas.0913008107

Gauthier, I., Skudlarski, P., Gore, J. C., & Anderson, A. W. (2000). Expertise for cars and birds recruits brain areas involved in face recognition. *Nat Neurosci, 3*(2), 191-197. doi:10.1038/72140

Gauthier, I., Tarr, M. J., Anderson, A. W., Skudlarski, P., & Gore, J. C. (1999). Activation of the middle fusiform 'face area' increases with expertise in recognizing novel objects. *Nat Neurosci, 2*(6), 568-573.

Guedel, A. E. (1937). Inhalation Anesthesia: A Fundamental Guide. *Anesthesia & Analgesia, 16*(2), 119-120.

Haun, A. M., Oizumi, M., Kovach, C. K., Kawasaki, H., Oya, H., Howard, M. A., . . . Tsuchiya, N. (2016). Contents of Consciousness Investigated as Integrated Information in Direct Human Brain Recordings. *bioRxiv*, 039032. doi:http://dx.doi.org/10.1101/039032

Healy, K., McNally, L., Ruxton, G. D., Cooper, N., & Jackson, A. L. (2013). Metabolic rate and body size are linked with perception of temporal information. *Animal Behaviour, 86*(4), 685-696. doi:https://doi.org/10.1016/j.anbehav.2013.06.018

Hendricks, J. C., Finn, S. M., Panckeri, K. A., Chavkin, J., Williams, J. A., Sehgal, A., & Pack, A. I. (2000). Rest in Drosophila is a sleep-like state. *Neuron, 25*(1), 129-138.

Herculano-Houzel, S., Mota, B., & Lent, R. (2006). Cellular scaling rules for rodent brains. *Proceedings of the National Academy of Sciences, 103*(32), 12138-12143. doi:10.1073/pnas.0604911103

King, J.-R., Sitt, Jacobo D., Faugeras, F., Rohaut, B., El Karoui, I., Cohen, L., . . . Dehaene, S. (2013). Information Sharing in the Brain Indexes Consciousness in Noncommunicative Patients. *Current Biology, 23*(19), 1914-1919. doi:10.1016/j.cub.2013.07.075

Koch, C., Massimini, M., Boly, M., & Tononi, G. (2016). Neural correlates of consciousness: progress and problems. *Nat Rev Neurosci, 17*(5), 307-321. doi:10.1038/nrn.2016.22

Kriegel, U. (2007). A cross-order integration hypothesis for the neural correlate of consciousness. *Consciousness and Cognition, 16*(4), 897-912. doi:http://dx.doi.org/10.1016/j.concog.2007.02.001

Ku, S.-W., Lee, U., Noh, G.-J., Jun, I.-G., & Mashour, G. A. (2011). Preferential Inhibition of Frontal-to-Parietal Feedback Connectivity Is a Neurophysiologic Correlate of General Anesthesia in Surgical Patients. *PLOS ONE, 6*(10), e25155. doi:10.1371/journal.pone.0025155

Lamme, V. A. F. (2010). How neuroscience will change our view on consciousness. *Cognitive Neuroscience, 1*(3), 204-220. doi:10.1080/17588921003731586

Lee, U., Mashour, G. A., Kim, S., Noh, G.-J., & Choi, B.-M. (2009). Propofol induction reduces the capacity for neural information integration: Implications for the mechanism of consciousness and general anesthesia. *Consciousness and Cognition, 18*(1), 56-64. doi:http://doi.org/10.1016/j.concog.2008.10.005

Liu, W. H. D., Thorp, T. A. S., Graham, S. G., & Aitkenhead, A. R. (1991). Incidence of awareness with recall during general anaesthesia. *Anaesthesia, 46*(6), 435-437. doi:10.1111/j.1365-2044.1991.tb11677.x

Llinás, R., & Ribary, U. (1993). Coherent 40-Hz oscillation characterizes dream state in humans. *Proceedings of the National Academy of Sciences, 90*(5), 2078-2081. doi:10.1073/pnas.90.5.2078

Massimini, M., Ferrarelli, F., Murphy, M. J., Huber, R., Riedner, B. A., Casarotto, S., & Tononi, G. (2010). Cortical reactivity and effective connectivity during REM sleep in humans. *Cognitive Neuroscience, 1*(3), 176-183. doi:10.1080/17588921003731578

Mather, J. A. (2008). Cephalopod consciousness: behavioural evidence. *Consciousness and Cognition, 17*(1), 37-48.

Mayner, W., Marshall, W., & Marchman, B. (2016). pyphi: 0.8.1. Zenodo. http://doi.org/10.5281.zenodo.55692

Mitra, P., & Bokil, H. (2007). Observed brain dynamics. New York: Oxford UP.

Morin, A. (2006). Levels of consciousness and self-awareness: A comparison and integration of various neurocognitive views. *Consciousness and Cognition, 15*(2), 358-371. doi:http://dx.doi.org/10.1016/j.concog.2005.09.006

Oizumi, M., Albantakis, L., & Tononi, G. (2014). From the phenomenology to the mechanisms of consciousness: integrated information theory 3.0. *PLoS Comput Biol, 10*(5), e1003588.

Oizumi, M., Amari, S.-i., Yanagawa, T., Fujii, N., & Tsuchiya, N. (2016). Measuring Integrated Information from the Decoding Perspective. *PLOS Computational Biology, 12*(1), e1004654. doi:10.1371/journal.pcbi.1004654

Paulk, A. C., Zhou, Y., Stratton, P., Liu, L., & van Swinderen, B. (2013). Multichannel brain recordings in behaving Drosophila reveal oscillatory activity and local coherence in response to sensory stimulation and circuit activation. *Journal of Neurophysiology, 110*(7), 1703-1721. doi:10.1152/jn.00414.2013

Person, A. L., & Raman, I. M. (2012). Synchrony and neural coding in cerebellar circuits. *Frontiers in Neural Circuits, 6*, 97. doi:10.3389/fncir.2012.00097

Pierce, K., Müller, R. A., Ambrose, J., Allen, G., & Courchesne, E. (2001). Face processing occurs outside the fusiform `face area' in autism: evidence from functional MRI. *Brain, 124*(10), 2059-2073. doi:10.1093/brain/124.10.2059

Rani, D. D., & Harsoor, S. S. (2012). Depth of general anaesthesia monitors. *Indian Journal of Anaesthesia, 56*(5), 437-441. doi:10.4103/0019-5049.103956

Rényi, A. (1961). *On measures of entropy and information.* Paper presented at the Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics.

Ross, E. D., Thompson, R. D., & Yenkosky, J. (1997). Lateralization of affective prosody in brain and the callosal integration of hemispheric language functions. *Brain and language, 56*(1), 27-54.

Rubner, Y., Tomasi, C., & Guibas, L. J. (2000). The earth mover's distance as a metric for image retrieval. *International journal of computer vision, 40*(2), 99-121.

Sareen, P., Wolf, R., & Heisenberg, M. (2011). Attracting the attention of a fly. *Proceedings of the National Academy of Sciences, 108*(17), 7230-7235. doi:10.1073/pnas.1102522108

Sebel, P. S., Bowdle, T. A., Ghoneim, M. M., Rampil, I. J., Padilla, R. E., Gan, T. J., & Domino, K. B. (2004). The Incidence of Awareness During Anesthesia: A Multicenter United States Study. *Anesthesia & Analgesia, 99*(3), 833-839. doi:10.1213/01.ane.0000130261.90896.6c

Seelig, J. D., & Jayaraman, V. (2015). Neural dynamics for landmark orientation and angular path integration. *Nature, 521*(7551), 186-191. doi:10.1038/nature14446

Sergent, C., & Dehaene, S. (2004). Neural processes underlying conscious perception: experimental findings and a global neuronal workspace framework. *Journal of Physiology-Paris, 98*(4), 374-384.

Shaw, P. J., Cirelli, C., Greenspan, R. J., & Tononi, G. (2000). Correlates of sleep and waking in Drosophila melanogaster. *Science, 287*(5459), 1834-1837.

Toker, D., & Sommer, F. (2016). Moving Past the Minimum Information Partition: How To Quickly and Accurately Calculate Integrated Information. *arXiv preprint arXiv:1605.01096*.

Toker, D., & Sommer, F. T. (2017). Great Than The Sum: Integrated Information In Large Brain Networks. *ArXiv e-prints, 1708*. Retrieved from http://adsabs.harvard.edu/abs/2017arXiv170802967T

Tononi, G. (2008). Consciousness as Integrated Information: a Provisional Manifesto. *The Biological Bulletin, 215*(3), 216-242. doi:10.2307/25470707

van Alphen, B., Yap, M. H. W., Kirszenblat, L., Kottler, B., & van Swinderen, B. (2013). A Dynamic Deep Sleep Stage in Drosophila. *The Journal of Neuroscience, 33*(16), 6917.

van Swinderen, B. (2005). *The Remote Roots of Consciousness in Fruit-fly Selective Attention?* (Vol. 27).

van Swinderen, B. (2006). A succession of anesthetic endpoints in the Drosophila brain. *Journal of Neurobiology, 66*(11), 1195-1211. doi:10.1002/neu.20300

Witter, L., Rudolph, S., Pressler, R. T., Lahlaf, S. I., & Regehr, W. G. (2016). Purkinje Cell Collaterals Enable Output Signals from the Cerebellar Cortex to Feed Back to Purkinje Cells and Interneurons. *Neuron, 91*(2), 312-319. doi:10.1016/j.neuron.2016.05.037

Yu, F., Jiang, Q.-j., Sun, X.-y., & Zhang, R.-w. (2015). A new case of complete primary cerebellar agenesis: clinical and imaging findings in a living patient. *Brain, 138*(6), e353-e353. doi:10.1093/brain/awu239