**Consciousness and structure of information: integrated information structure collapses with anesthetic loss of conscious arousal in Drosophila melanogaster**

Angus Leung1,\*, Dror Cohen1,2, Bruno van Swinderen3, Naotsugu Tsuchiya1,2,4,5,\*

1 School of Psychological Sciences, Monash University

2 Center for Information and Neural Networks (CiNet), National Institute of Information and Communications Technology (NICT), Osaka

3 Queensland Brain Institute, The University of Queensland

4 Monash Institute of Cognitive and Clinical Neuroscience (MICCN), Monash University

5. Advanced Telecommunications Research Computational Neuroscience Laboratories, 2-2-2 Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-0288, Japan.

Corresponding authors: AL, NT

Nature Neuroscience format:

Abstract 150 words

Intro > Results > Discussion 4500 words

Methods

250 words per figure legend

Replace

small-phi - φ

big-phi - Φ - system-level II

II - Φ

IIS - compositional structure

Purview - fine as is (relatively few)

Mechanism

Order - size; number of channels/mechanisms (if word limit allows)

unpartitioned/partitioned - full/disconnected

**Abstract**

The physical basis of consciousness remains one of the most elusive concepts in current science. One influential conjecture is that consciousness is to do with some form of information or causality. The integrated information theory (IIT) of consciousness proposes that a high-level consciousness with rich and specific contents corresponds directly to a hierarchically organised, irreducible pattern of causal interactions, that is, an informational structure among elements in a system. Here, we tested this conjecture in a simple biological system (fruit flies), estimating the information structure of the system during wakeful and anesthetized conditions. We found that patterns of causal interactions among populations of neurons in the fly brains during wakefulness collapsed during isoflurane anesthesia, rendering them to isolated clusters of interactions, as demonstrated by the index of consciousness proposed by IIT. We used classification analysis to quantify the accuracy of discrimination between conscious states; we found that inforamtional structures inferred from more channels discriminates conscious states better. Further, we found that the collapse allowed for better discrimination of wake and anesthesia in more central regions of the brain. Our results speaks a potential utility of the novel concept of “informational structure” as a measure for the level of consciousness, which in turn provides a hitherto missing link between level and contents of consciousness; level and contents of consciousness corresponds to the volume and the specific shape of the information structure.

# Introduction

The question of how subjective, conscious experience arises from physical interactions has been pondered by philosophers for centuries (e.g. Nagel 1974 What is it like to be a bat? Chalmers 1995 conscious mind; need very old reference to back “centuries” claim), and now has moved into the domain of cognitive neuroscience (Dahaene et al. 2014 book, Lamme 2015 Crack of Dawn, Koch 2016 Nat Rev Neuro). Because we are only able to experience our own individual consciousness, exact inference of others conscious contents (i.e., what it is like to be a bat [Nagel 1974]) seem intractable. However, more broader and looser inference on levels of consciousness (i.e., whether any biological organism is consciously wakeful or not) seems reasonably well inferred based on a multitude of relationships among one’s own experience across varied levels of consciousness, anatomical and physiological differences in brains, and associated functions and behaviors (Velmans 2009 J Conscious Studies, Tegmark 2017 Life 3.0, Tsuchiya & Saigo (in prep)). Indeed, such inferences have been widely accepted across various loss of consciousness in brain damaged patients [Laureys 2015 Neurology of Consciousness] and animals [Hudetz Mashour 2016 Anesthesia - anesthesia on monkeys, Yanagawa 2013 PLoS One - Monkey ECoG], and now becoming possible to apply to insects [Orecia & Bruno’s review, Dror’s papers, and other insect anesthesia papers, Sleep Shaw2000 Science ].

One influential view is that consciousness arises from “self-existence”, as measured through “intrinsic information” (II). This view comes from both philosophy (Chalmers 1995) and physics (Tegmark 2017 Life 3.0). With this idea, integrated information theory (IIT; Oizumi 2014; Tononi 2016) proposes a highly specific, operationalised concept of intrinsic information as critical for the arisal of consciousness. II, distinct from the standard notion of Shannon information (cite Shannon paper 1948 or textbook), is defined as “differences that make a difference within a system” (Bateson 1972, Tononi 2008, Oizumi 2014). In other words, II is concerned with how elements of a system causally influence each other such that information is accessible to the system itself (extrinsic information, conversely, concerns how states of a system casually influence states of another, separate system; see supporting information in Oizumi 2016 PLoS Comp).

IIT provides a mathematical formulation of II which allows us to quantify and characterise how it is structurally organised and composed within a system. More specifically, IIT characterises how hierarchically organised elements uniquely and causally interact with other elements within a system in an irreducible (i.e. integrated) manner. The derived intrinsic information structure (IIS; i.e., maximally irreducible conceptual structure; Oizumi 2014) is hypothesised to directly correspond to the quantity and quality of consciousness. That is, the richer and more specific the informational structure of the system, the higher the level of consciousness in a system, and the richer the contents that the system consciously experiences. Importantly, IIT formalises the IIS in a manner which does not assume a specific biology (e.g. the human brain), and can in principle be applied in any system. .

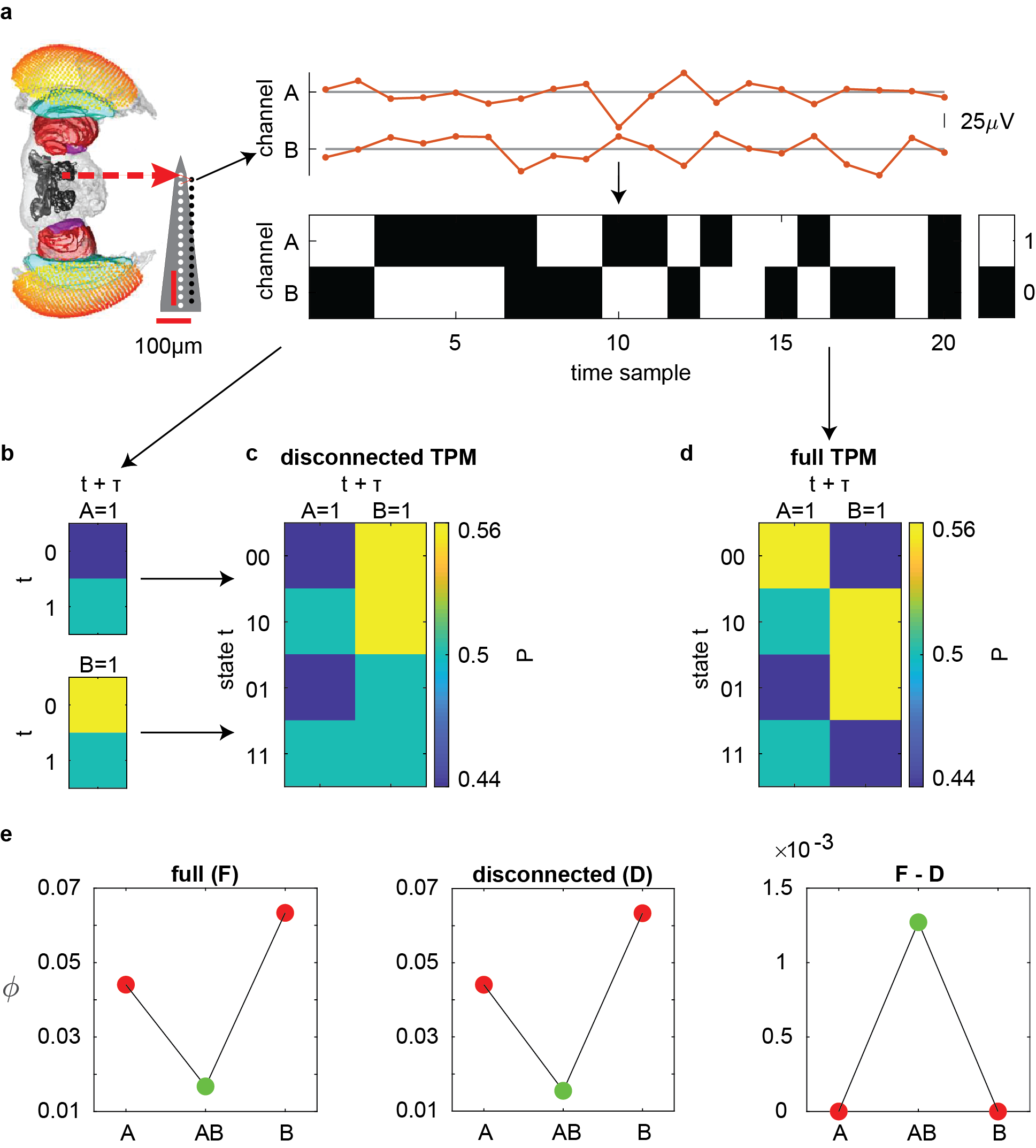
Given this theoretical background, we specifically address the following questions. Firstly, how can we apply the IIT formalism to neural activity recorded from a biological system in order to estimate its intrinsic information structure (IIS)? Secondly, does the estimated IIS shrink when animals are anesthetised compared to when they are awake? And thirdly, can estimating the IIS give us any biological hints as to the “interesting” regions of the brain to look for consciousness? To address these questions, we use neural recordings from a simple biological system (fruit flies), from which we obtained 15 locally referenced population neural activities during wakefulness and during anesthesia (Cohen 2017, 2018). By developing a procedure to estimate the IIS of the system from these recordings using an existing toolbox [Mayner 2018 PLoS Comp], we were able to compute IIS from neural recording, finding that the IIS is indeed reduced during anesthesia, and that it is more reliably reduced in the central regions of the brain.

# Results

**Constructing intrinsic informational structures from fly local field potentials**

We computed integrated information structures (IIS) across sets of 15 local field potentials (LFPs; hereafter referred to as “channels”) from 13 flies across 8 epochs of each wakefulness and anesthesia (see Methods; Cohen 2018). LFPs were obtained from a linear multi-electrode array inserted laterally to the eye, such that the 15 channels cover roughly half the fly brain and record voltages from both peripheral and central regions of the brain. From these 15 channels, we considered every possible combination of 4 channels as a system, giving 1365 channel sets per fly.

To construct the IIS, we first needed to operationalise the “state” of each set of channels by discretizing LFP voltages into binary values (Fig. 1a; for illustrative purposes we show the process for a subset of 2 channels). We binarised voltages by taking the median voltage for each channel across all time-samples within a trial, and then converting each time-sample in the trial to ‘on’ if the voltage for that time-sample was greater than the median, and ‘off’ otherwise.



**Fig. 1: Processing pipeline for computing the IIS. a,** LFPs are recorded from the fly brain using a linear multi-electrode probe. Re-referenced LFPs are discretized by comparing to the median voltage for each trial. Displayed is an example of 20 samples for a set of two channels. The network state at a given sample is given by the state of its channels - either on (above the median LFP for the channel) or off (equal to or below the median). **b,** After discretisation, a transition probability matrix (TPM) for each channel describes how its state transitions from time t to time t+τ (here τ=4ms). Each cell in the TPM gives the probability of each channel in the full network transitioning to state ‘1’ given the state of the network in the past. A TPM can be calculated for A and B individually, treating the system as independent parts. These probabilities are obtained from the full TPM in d. **c,** Based on TPMs for each part, we can obtain a disconnected TPM of the network, which assumes independence between the parts. **d,** The full TPM is constructed by considering the two channels simultaneously, and represents the causal interaction of the full system. **e,** At a given state (e.g. 00), we find the integrated information II contributed by each subgroup of channels (leftmost panel) from the full set by comparing actual transition probabilities associated with the state to those associated after “splitting” the subgroup’s connections (full structure; F). This is repeated after specifying a cut at the system level, yielding a disconnected structure (D). As there are multiple possible cuts, we take D which is most similar to F, thus filtering for information which only exists when the system is whole. System level II (SII) is the distance between all concepts in UP with those in P. Here we summarise the distance as the difference in II values between corresponding concepts (F-D). As information is assessed by comparing probability distributions, if the disconnected TPM is identical to the full TPM, SII is 0. A more detailed explanation of the process is described in Table 1.

Using the time series of binarised states, we then computed the probabilities of each state transitioning into any other state, constructing a transition probability matrix (TPM; Fig. 1b, c and d). Specifically, we calculated transition probabilities for each channel being “on” at time *t* given the state of all channels at *t* - 𝜏. These transition probabilities describe how the set of channels changes over time, with every state being associated with a probability distribution of possible past and future states. By comparing the distributions specified by the TPM to maximum entropy distributions, we measure the information generated by the state from the system’s perspective (note that the notion of information in IIT refers to “differences that make a difference”, and is distinct from the standard notion of information in the context of Shannon information theory; see Supplementary material in Oizumi 2014). We then constructed minimally reducible versions of the TPM by assuming causal independence among parts (disconnecting the system; see Methods). By comparing actual transition probabilities to reduced transition probabilities, we assess the degree to which a group of subsystems generate information above and beyond their independent parts, i.e. integrated information (II).

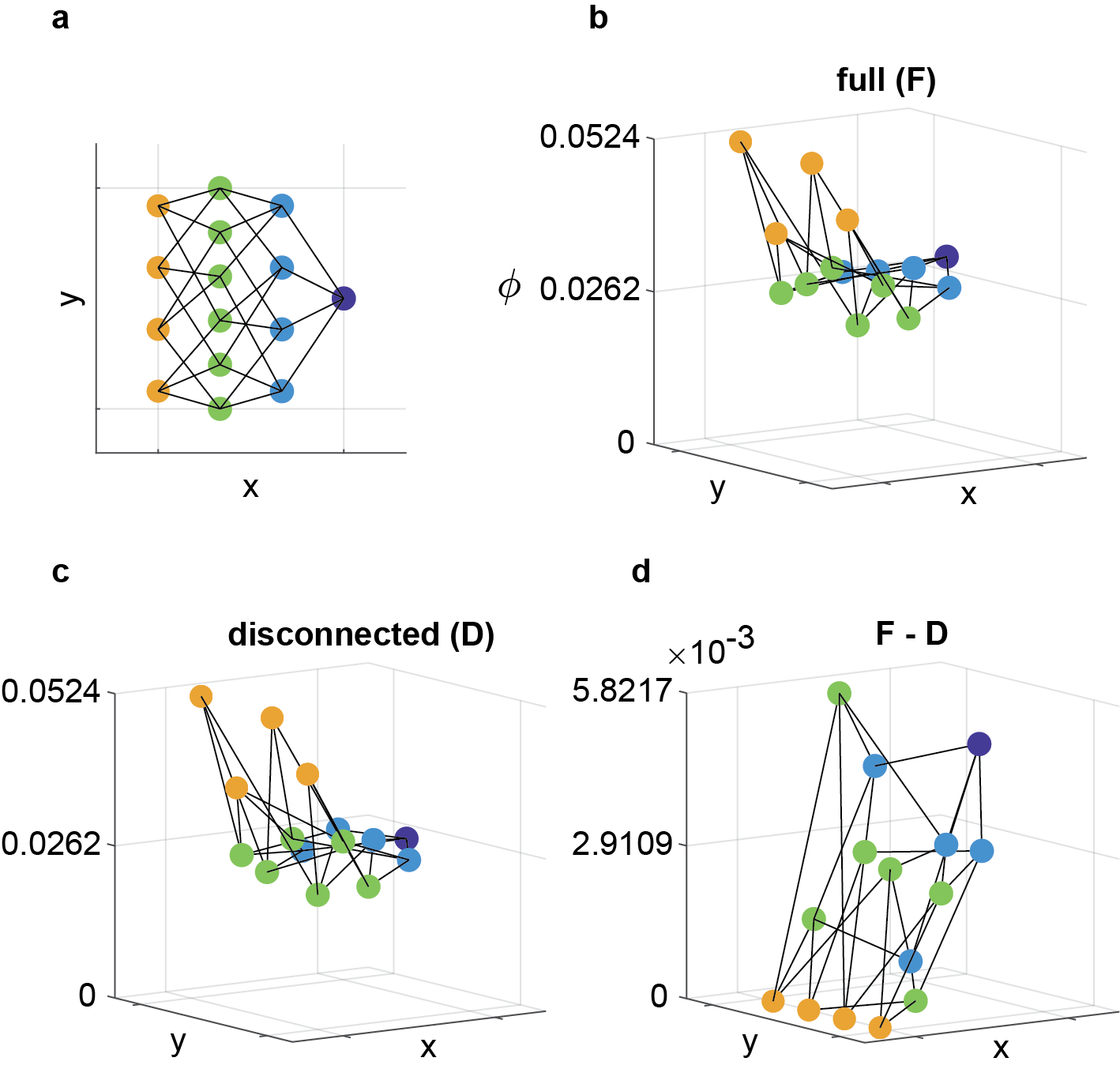
Fig. 1e shows the II generated for the full two-channel system by each subset of channels. II of a candidate subset (“mechanism”) is quantified with respect to subsets of channels (“purviews”) which either affect or are affected by the candidate subset. Note that mechanisms with more channels are composed of mechanisms with fewer channels (e.g. AB is composed of A and B). Table 1 lists a basic algorithmic flow of determining which purview to use when determining the amount of II generated by a given mechanism. Summarily, for a given mechanism, a) II is computed across all purviews, b) at a given purview, II is obtained by finding the disconnection which isolates irreducible causes or effects (and consequently minimises II), and c) take the maximum II across purviews. This is repeated for both causes and effects, and the minimum II out of the two is taken as the maximally irreducible difference that the mechanism makes to the system, and as the II generated by the mechanism for the system.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Candidate cause | Partition | Partition (new format) | II (degree of difference the mechanism makes) | Within a candidate cause, does the partition minimise II? |
| A | AB/[ ] x [ ]/A | AB -> [ ]  -----------  [ ] -> A | 0.0426 | N |
| A | A/[ ] x B/A | A -> [ ]  -----------  B -> A | 0.0441 | N |
| A | A/A x B/[ ] | A -> A  -----------  B -> [ ] | 0.0015 | Y |
| B | AB/[ ] x [ ]/B | AB -> [ ]  -----------  [ ] -> B | 0.0480 | N |
| B | A/[ ] x B/B | A -> [ ]  -----------  B -> B | 0.0154 | Y |
| B | A/B x B/[ ] | A -> B  -----------  B -> [ ] | 0.0635 | N |
| AB | [ ]/A x AB/B | AB -> B  -----------  [ ] -> A | 0.0427 | N |
| AB | [ ]/B x AB/A | AB -> A  -----------  [ ] -> B | 0.0481 | N |
| AB | [ ]/AB x AB/[ ] | AB -> [ ]  -----------  [] -> AB | 0.0907 | N |
| AB | A/[ ] x B/AB | A -> [ ]  -----------  B -> AB | 0.0595 | N |
| AB | A/A x B/B | A -> A  -----------  B -> B | 0.0167 | Y |
| AB | A/B x B/A | A -> B  -----------  B -> A | 0.1078 | N |
| AB | A/AB x B/[ ] | A -> AB  -----------  B -> [ ] | 0.0650 | N |

**Table 1: Determination of mechanism integrated information (II).** II of a mechanism (here we use a 2-channel mechanism AB from a 2-channel system {A, B}) is determined through finding the maximal cause and effect (for simplicity, we show only cause) for the mechanism. To determine the cause we search through all subsets of channels in the system (A, B, and AB, highlighted in green, yellow, and blue respectively). For each candidate cause we measure the extent to which the mechanism AB, above and beyond its constituent parts A and B, was caused by the candidate. We measure this as the distance between the actual probabilities of possible past states of the cause conditioned on AB, and the reduced probabilities conditioned on A and B separately, i.e. II. As there are multiple ways to disconnect AB in relation to the candidate cause, we take the disconnection which gives the smallest distance, with this distance reflecting the information specified by AB which is completely unaccounted for by A and B independently. Finally, we take the cause for which II is maximal. This process is repeated for the possible effects of AB, and the minimum of cause and effect is taken as the II generated by the mechanism.

The whole process of constructing the IIS is repeated after disconnecting the full system (see Methods for details). In short, we consider all possible disconnected model systems, where a subset of connections statistically disconnected. From the disconnected models, we select the one which best approximates the IIS of the full system. In this manner, two IISs are obtained, one corresponding to the full, whole system, and another corresponding a disconnected system which best approximates the full system (Fig. 1e middle). System level integrated information (SII)[[1]](#footnote-0) is the sum of distances between the probability distributions specified by all mechanisms in the full system and the disconnected system. If the IISs are identical despite the disconnection, the system is deemed reducible into independent parts and generates SII = 0. To illustrate differences between the full and disconnected IISs, we show the difference in II values for all mechanisms (Fig. 1e, right).

Fig. 2 shows an example full and disconnected IIS obtained from 1 fly, 1 channel set during wakefulness, as well as the difference between the full and disconnected IISs, when extending this process to the 4-channel case. [Add a link to a 3D movie -> upload it to OSF, then make a link to it.]



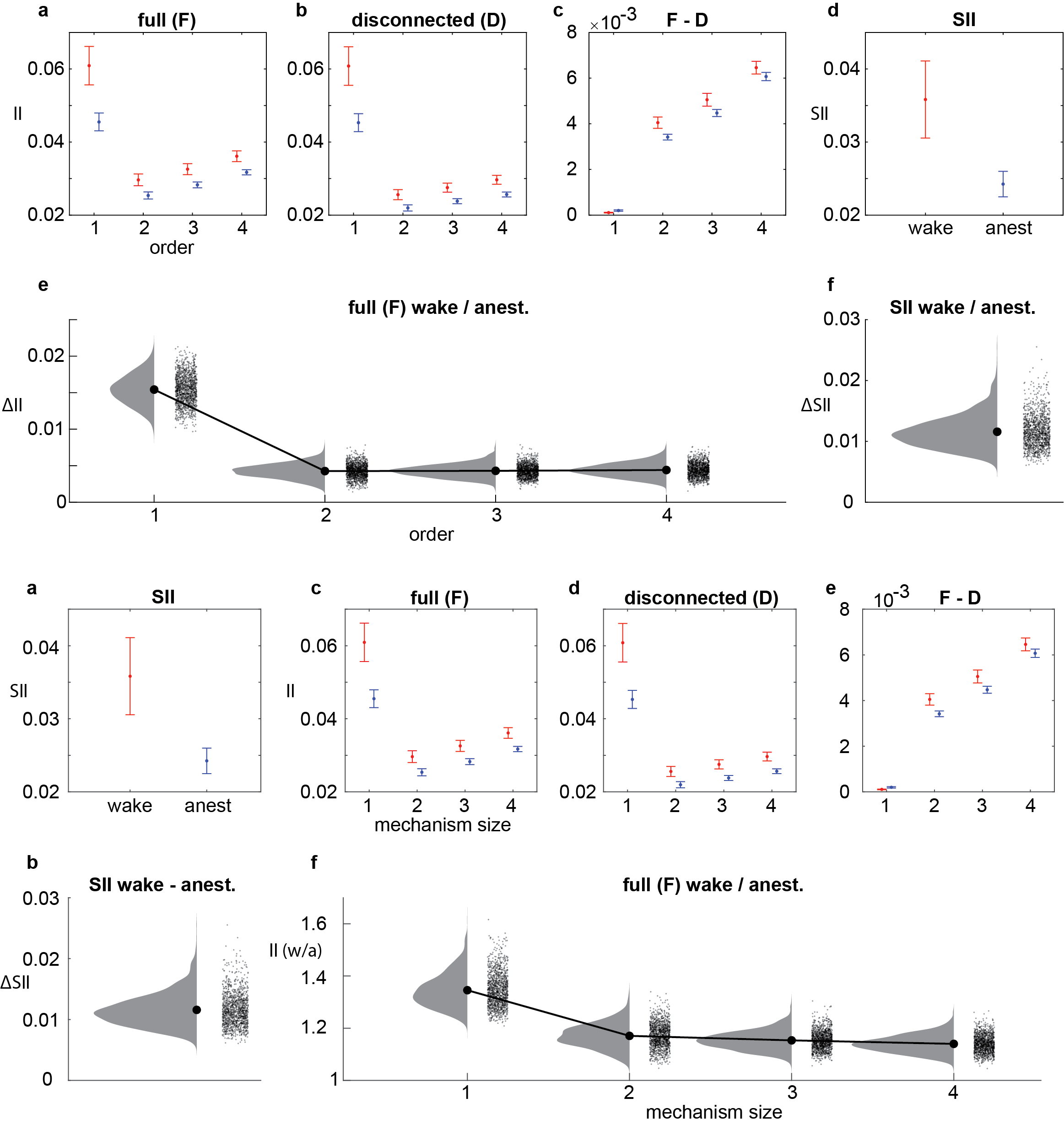
**Fig. 2: Integrated information structure (IIS) for one channel set for one fly in the awake state. a,** Top-down view of the IIS. The x-axis is chosen to convey the number of channels that constitute each mechanism (the “order”) (yellow, green, light blue, and dark blue dots indicate mechanisms consisting of 1, 2, 3, and 4 channels respectively). The y-axis is arbitrarily set to give equal spacing between mechanism. Lines indicate inclusion relations (e.g., mechanism AB consists of A and B). **b,** An exemplar disconnected IIS from a single fly, channel set, and trial. **c,** The corresponding IIS after imposing disconnections which best approximate the full IIS. **d,** The distance between the full and and disconnected IISs, SII, is the sum of distances, across all mechanisms, between probability distributions which specify for mechanism II before and after disconnecting the full system. For illustration, we show the difference in IIs between the full and disconnected IISs, for all mechanisms.

**Integrated informational structure collapses due to general anesthesia**

IIT proposes that the IIS is isomorphic to the structure of conscious experience of the system - when experience is reduced due to loss of consciousness, the IIS should collapse. The collapse of the IIS should be reflected in the II values of each mechanism, as well as the overall SII. Thus, we expected a) SII, which quantifies the difference between the full and disconnected systems’ IISs to be reduced by anesthesia and b) II for each mechanism to be reduced by anesthesia (opposed to e.g. increased II for some mechanisms, which would be an interesting outcome).

Further, SII cannot increase when higher order mechanisms are not integrated. Consider a case of two independent pairs of connected neurons. In such a case, II for 2-channel mechanisms could be high. However, if the two pairs are not connected at all, then II for the overall 4-channel mechanism would be zero. As anesthesia disrupts global communication, we thus reasoned that higher order mechanisms should reflect more of the overall collapse of the IIS. Given this, we expected that II for larger mechanisms consisting of more channels would be affected more than anesthesia than II of mechanisms with fewer channels.

We first checked the clearest prediction of IIT - that SII, IIT’s proposed measure of level of consciousness, should be reduced during anesthesia. Using linear mixed effects analysis (to account for intra-fly channel set correlations; see Methods), we found SII to significantly differ between conscious arousal levels (Fig. 3a,b; 𝜒2(3) = 6656, *p* < .001), being reduced during anesthesia (*t*(35488) = -85.57, *p* < .001).



**Fig. 3: Effect of anesthesia on the IIS. a,** SII values during wakefulness (red) and anesthesia (blue). Error bars indicate standard error across N=13 flies (SII and II values were first averaged within flies for a, c, d, and e). **b,** Difference in SII for all channel sets, averaged across flies. **c,** II values from the full IIS for each mechanism size, averaged within size and across channel sets and flies, for wake (red) and anesthesia (blue). **d,** II values as in (a), but for the disconnected IIS. **e,** Difference in II values between the full and disconnected IIS. **f,** Difference in II values between wakefulness and anesthesia (wakefulness - anesthesia for the full IIS. Here we show, for all 1365 channel sets, the average difference across concepts and flies at each concept order.

We next compared II for each mechanism obtained during wakefulness to those obtained during anesthesia. Using linear mixed effects analysis (to account for intra-fly channel set correlations; see Methods), we found II to be reduced during anesthesia (Fig. 3c,f; 𝜒2(1) = 3.092 × 104, *p* < .001). However, II varied with the size of the mechanism (𝜒2(3) = 1.512 × 105, *p* < .001), and the interaction between anesthesia and mechanism size was significant (𝜒2(3) = 1.203 × 104, *p* < .001), indicating that the extent to which II was reduced due to anesthesia varied across mechanism sizes.

To break down the significant interaction between mechanism order and conscious arousal level, we first describe the general shape of the IIS of the full system during wakefulness by comparing II among mechanism sizes (Fig. 3c; see supplementary materials for the disconnected IIS, Fig. 3d). Pairwise comparisons (see Methods) indicated that II generally increased with mechanism size, with 3-channel mechanisms having greater II than 2-channel mechanisms, and 4-channel mechanisms having greater II than 3-channel mechanisms (*t*(1.775 × 105) = 141.63 and *t*(88723) = 178.16, respectively, *p* < .001 for both), with the exception of the 1-channel mechanisms which overall had the greatest II (*t*(88723) = 95.10, compared to 4th order). This pattern was mirrored in the anesthetised condition, as well as the partitioned IIS (see Supplementary materials).

To test if higher order mechanisms were more greatly affected by anesthesia, we looked at the differential effects of anesthesia at each mechanism order in the full IIS. As the amount of II generated varied across mechanism sizes, we compared the ratio of wakeful to anesthetised II across mechanism sizes. A larger ratio corresponds to a larger decrease in II due to anesthesia. We verified that the ratio of wakeful to anesthetised II was also significantly different among mechanism sizes (𝜒2(3) = 2.229 × 104, *p* < .001). However, instead of finding higher order mechanisms to have larger relative reductions in II due to anesthesia, we found the opposite where larger mechanisms had smaller relative reductions (*p* < .001 for all comparisons, *t*(1.775 × 105) = -106.8, *t*(1.775 × 105) = -34.22, and *t*(88723) = -22.38, for comparing 1st to 2nd, 2nd to 3rd, and 3rd to 4th order mechanisms respectively).

We next looked at how the change in the IIS due to disconnecting the full model is affected by anesthesia. To do this, we compared the difference in II at each mechanism between the IIS for the full and disconnected system (delta-II) during wakefulness and anesthesia (Fig. 3e). delta-II reflects the degree to which disconnecting the full model affects a mechanism. As for the IIS for the full system, we found mechanism order and conscious arousal level to significantly affect delta-II (𝜒2(3) = 5.403 × 104, *p* < .001 and 𝜒2(1) = 2.365 × 102, *p* < .001 respectively), as well as their interaction (𝜒2(3) = 1.013 × 104, *p* < .001). Comparing mechanism sizes during wakefulness revealed that larger mechanisms had larger delta-II values (*t*(1.775 × 105) = 680.5, *t*(1.775 × 105) = 161.6, and *t*(88723) = 146.4, *p* < .001, for comparing 1- to 2-channel, 2- to 3-channel, and 3- to 4-channel mechanisms respectively), indicating that they are more greatly affected by the disconnecting of channel sets into smaller, independent sets. For 1-channel mechanisms, delta-II was often 0 (53% and 40% of all mechanisms during wakefulness and anesthesia). However, delta-II was non-zero for all other mechanism sizes. As for the IIS for the full system, the ratio of wake to anesthesia delta-II decreased with increasing mechanism size (overall effect of mechanism size, 𝜒2(3) = 569.7, *p* < .001; *t*(1.497 × 105) = -17.59, *t*(1.775 × 105) = -41.98, and *t*(88723) = -31.51, *p* < .001, for comparing 1- to 2-channel, 2- to 3-channel, and 3- to 4-channel mechanisms respectively). Thus, while larger mechanisms are more affected by disconnecting the system (i.e. they hold information which is not available in a disconnected system), it is the smaller mechanisms which are most proportionally affected by anesthesia.

Given that the IIS indeed collapsed during anesthesia across all mechanism orders, and II is reduced, we next sought to determine whether larger mechanisms give better discriminability of conscious level than smaller mechanisms. In other words, given that II is reduced during anesthesia, is this decrease more reliable for larger mechanisms? We also sought to determine whether considering the full IIS allows for better discrimination conscious level than just consideration of single mechanisms, i.e. is the pattern of II useful above and beyond considering independent II values in isolation? As IIT proposes SII as the measure of conscious level (whereas the IIS is a measure of experiential contents), we further compared this to the reliability of the decrease in SII, where we expected SII to achieve the best classification accuracy.

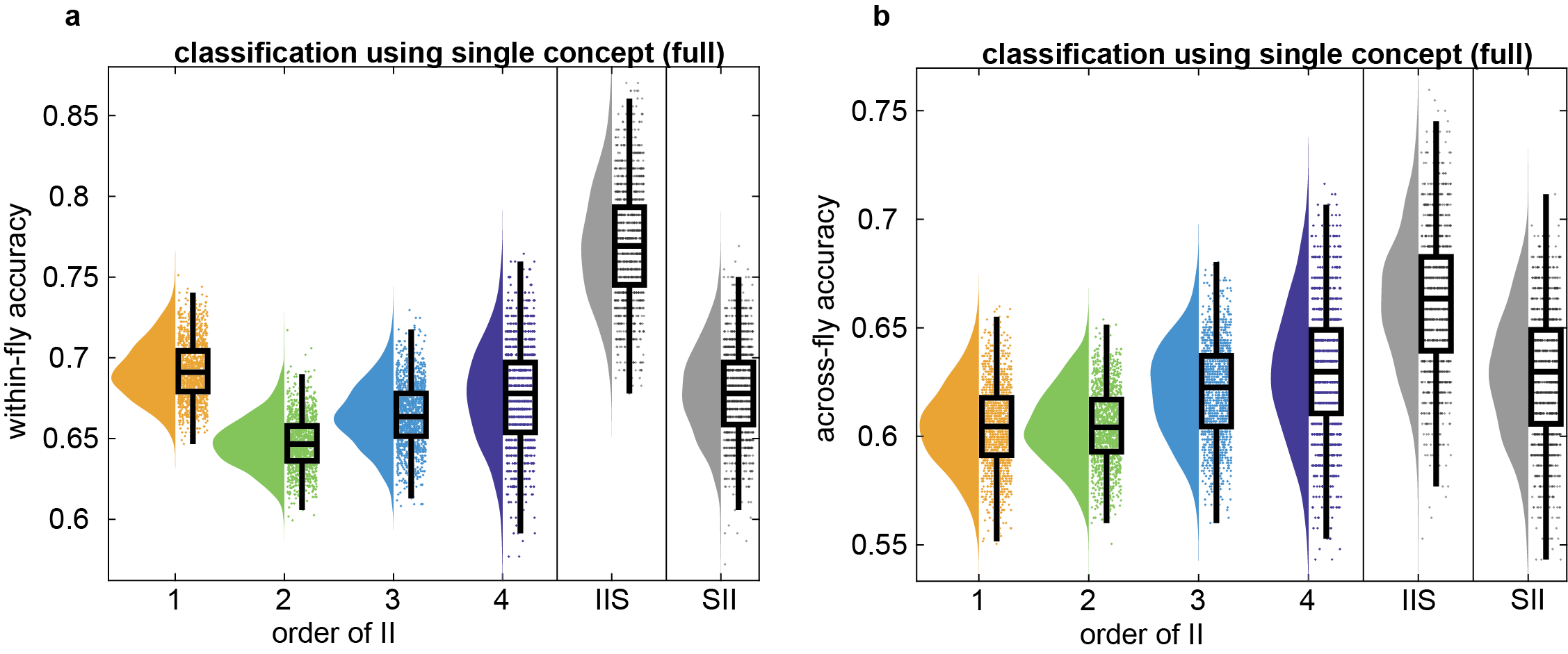
<Introduce both within and across fly classification together?>To compare the reliability of decreased II, SII, and the collapse of the full IIS, we employed classification analysis, which allows us to compare the reliability of one-dimensional changes of II and SII with multidimensional changes of the full IIS. We trained and tested support vector machines (SVMs) to classify conscious arousal level at two levels: a) classifying epochs within each fly (within-fly classification, repeated for each fly), and b) classifying flies at each trial (across-fly classification, repeated at each epoch). At both levels, we used leave-one-out cross-validation for each channel set, such that each validation the SVM was trained on 7 epochs of wake and anesthesia (within-fly classification) or 12 flies (across-fly classification), with the remaining epoch-pair or fly used for testing (see Methods).

<Start first with within-fly classification, then shift to across-fly classification>To compare the reliability of decreased II, SII, and the collapse of the full IIS, we employed classification analysis, which allows us to compare the reliability of one-dimensional changes of II and SII with multidimensional changes of the full IIS. We trained and tested linear support vector machines (SVMs; cite liblinear) to classify conscious arousal of epochs with each fly (within-fly classification), using leave-one-pair-out cross-validation for each channel set, such that at each validation the SVM was trained on 7 paired epochs of wakefulness and anesthesia, with the remaining epoch pair used for testing. We report the highest accuracies obtained at the optimal cost hyperparameter of the SVMs, determined independently for each channel set and fly, and for each measure used (mechanism II, SII, or the SII; see Methods).

<within-fly classification, bonferroni correction for post-hocs>We found that the majority of channel sets were able to discriminate wakeful from anesthesia with better than chance performance, regardless of whether II or SII was used (Fig. 4a). Further, classification accuracy varied significantly depending on what measure was used (𝜒2(5) = 1.563 × 104, *p* < .001). Unexpectedly, SII did not achieve the greatest classification accuracy, and actually performed significantly worse than II of 1-channel and 4-channel mechanisms (*t*(6823) = -11.11, *p* < .001 and *t*(2728) = 3.051, *p* = .035). Among independent mechanisms, 1-channel mechanisms on averaged classified wake from anesthesia with the highest accuracy (*t*(13648) = 74.24, *t*(10918) = 41.10, *t*(6823) = 13.51, compared to 2-channel, 3-channel, and 4-channel mechanisms respectively, *p* < .001 for all comparisons), consistent with them having the largest decrease in II due to anesthesia. However, 2- and 3-channel mechanisms, which had larger relative decreases in II due to anesthesia, performed worse than 4-channel mechanisms ().

<across-fly classification intro>To compare the reliability of decreased II, SII, and the full IIS, we employed classification analysis, which allows us to compare the reliability of one-dimensional changes of II and SII with multidimensional changes of the full IIS. We trained and tested support vector machines (SVMs) using leave-one-pair-out cross-validation for each channel set, such that the SVM was trained at each validation on 12 flies, with the remaining fly used for testing (see Methods). We emphasise that by doing this, we are classifying the conscious arousal of a given fly based on the observed values of other flies.

<across-fly classification, bonferroni correction for post-hocs - need to update stats>We found that the majority of channel sets were able to discriminate wakeful from anesthesia with better than chance performance, regardless of whether II or SII was used (Fig. 4). Further, classification accuracy varied significantly depending on what measure was used (𝜒2(5) = 2667, *p* < .001). Though we previously observed II to be less reduced during anesthesia for higher order mechanisms, here we found them to be more reliable, as reflected through higher classification accuracy (*t*(13648) = 13.2, *t*(13648) = 26.77, *t*(6823) = 8.093, comparing 1- to 2-channel, 2- to 3-channel, and 3- to 4-channel mechanisms respectively, *p* < .001 for all comparisons). Interestingly, accuracy achieved by SII was not significantly different to four-channel II (*t*(2728) = -2.16, *p* = .031). However, consideration of the full IIS gave on average the best accuracy overall (*t*(2728 = 5.240, *p* < .001, compared to 4-channel II).



**Fig. 4: Classification of wake/anest using each concept order.** Across-fly classification at each individual channel set using each concept order (single-feature classification). Individual points are classification accuracy of each channel set, after averaging accuracies across all mechanisms within the same mechanism order. The grey distributions show accuracy when using all mechanisms together (15-feature classification; left), and when using Φ3.0 (right). Boxplots show median, 25th-75th percentiles, and whiskers are 1.5 interquartile below and above respectively.

**Intrinsic information structure better distinguishes conscious level at central regions**

[Do we want to go into specific channel locations? Channels 1, 5, 10 seem to give good discriminability in general (see comment on Fig. 4)]

Having determined that the II and IIS are able to discriminate conscious level, we next asked whether they can be used in the context of finding brain regions which have an important role in holding conscious experience. Specifically, it is suspected that central regions of the fly brain are responsible for integrating feedforward inputs from peripheral regions (is it true? cite). Thus, we expected channel sets which are more centrally located (i.e. sets which consist of channels which are on average more centrally located) to have either a greater reduction in II or to have better discrimination of conscious level.

To test this, we computed correlations between big-phi with the average location channels (1 being most central, 15 being most peripheral). However, this yielded a weak positive correlation <TODO: double check this; stats>. Despite this, correlation between channel set location and classification accuracy of wake and anesthesia when using either the full IIS or big-phi was significantly negative [IIS: r = -0.192, p <.001; big-phi: r = -0.292, p < .001; 4th-order small-phi: r = -0.275, p < .001]

# Discussion

Here, we addressed two main points with regards to IIT 3.0. Firstly, we demonstrated the construction of an IIS as put forward by IIT from local field potentials recorded with multi-array electrodes from *Drosophila melanogaster* (fruit fly). Secondly, by inducing isoflurane anesthesia, we found that, as predicted by IIT, the IIS collapsed during anesthesia, accompanied by a reduction in integrated information Φ. In addition to these, we further found that the collapse of the IIS during anesthesia is more reliable across flies at more central regions of the fly brain.

IIT stands out as a theory of consciousness because it provides what is in essence a mathematical formulation of consciousness from first principles - by first identifying core aspects of conscious experience, and then deriving a measure of consciousness, it stands out by not only by taking an approach to tackling the question of how consciousness arises from physical interactions (starting from introspection of experience itself, instead of from empirical observations which are likely related to consciousness), but also by providing an operationalisation of its ideas, in the form of the IIS and Φ. Thus far, research on testing IIT has largely focused on Φ itself both using simulation of artificial [phi1.0: phi2.0, phi3.0 Albantakis, cite a lot of papers here] and biological systems [phi1.0: phi2.0, phi3.0 - Niisato arxiv 2019] and neglected the idea of the IIS (except our previous work; Haun et al 2017 eNeuro). This may be because it is not described in IIT as being directly linked to conscious level. However, we show here that not only is the IIS linked to conscious level, but it can possibly describe conscious level in a system better than Φ itself.

[discuss IIS shape, reason behind the shape, reason behind differential effect of anesthesia at different orders]

Our linking of the IIS with conscious level has two clear implications for IIT. Firstly, in the context of the computational complexity of big-phi, heuristics of big-phi which try to capture the core IIT ideas of information and integration (Nilsen 2019 Entropy) tend to focus only on approximating big-phi, with little regard for composition. Our results suggest that this is perhaps a naive approach, and that any computational approximations or derivations of big-phi should ideally also consider how to build up an IIS. Secondly, within a more theoretical/philosophical context, our results raise the question of exactly what “level” of consciousness means. Given that the IIS is able to discriminate wakefulness from anesthesia with greater accuracy than IIT’s own proposal of a measure of conscious “level”, exactly what a scalar measurement like big-phi contributes beyond the IIS is perhaps unclear. It may be the case that scalar values are simply not a sufficient summary of the multi-dimensional nature of conscious experience (Bayne?).

<Location effects>

Thirdly, characterising Φ3.0 across subsets of LFPs revealed an interesting pattern of Φ3.0 across the brain. Interestingly, Φ3.0 was not greatest just in channel sets containing central channels, but also across spatially close peripheral channels. This local peripheral increase of Φ3.0 may be reflecting the existence of both feedforward and feedback connections between the lamina and medulla in visual processing (Meinertzhagen 1991 Journal of Comparative Neurology, read in intro of Tuthill 2013 Neuron; Morante 2008 Curr Biol). Meanwhile, spatially close channels in the center may reflect similar connectivity in the midbrain (). Interestingly, Φ3.0 was larger not only in sets containing central channels, but also in spatially distributed sets which included central channels, suggesting that at the scale of the half-brain, regions are integrated informatively in that they, together, influence their activity. Previously, dynamics of effective connectivity have been characterised using bivariate measures. For example, anesthesia reduces feedback connectivity from central regions to the periphery, quantified with low-frequency component Granger Causality (Dhamala 2008 PRN, Wen 2013 Philos Trans), while leaving feedforward connectivity, quantified using high-frequency component Granger Causality, unaltered. As Φ3.0 requires both feedback and feedforward causality to be high, these results are reflected in a measure designed to be applied in a multivariate context. Given that reduction of feedback connectivity due to anesthesia from frontal to posterior brain regions have been reported in human EEG studies {Lee 2009, 2013, Boly 2012, Jordan 2013, Ranft 2016 -- See Cohen 2018 eNeuro ref}, it is likely that the sensitivity of Φ3.0 extends beyond the fly model.

Given that Φ3.0 is altered due to anesthesia, it was expected that it could be used to classify wake/anesthesia. Interestingly, we did not find optimal classification to occur at those channel sets which exhibited the largest change in Φ3.0 (those with spatially close channels at the periphery), but instead at those between the center and periphery, corresponding roughly to the lobula, the higher-order visual processing center of brain. This suggests that the bridging between peripheral and central regions is most consistently affected by anaesthesia. Past studies have reported similar effects, regarding loss of interaction between peripheral and central regions, but without characterising the spatial dimension in a continuous manner (Cohen 2018). The pattern of channel sets with better performance was interestingly accentuated when classifying across flies, with across-fly classification performing overall better than within-fly classification. Better performance for across-fly performance may be reflecting the manner in which Φ3.0 was derived, from core principles instead of from observing specific interactions and linking them back to consciousness.

Existing literature suggests that processing related to arousal takes place at a timescale of roughly 20-30 Hz (van Swinderen Curr Bio 2004). However using big-phi, we show that we can already distinguish between conscious arousal levels at a much shorter timescale, 4 ms.

We acknowledge several theoretical caveats. In particular, while we tested IIT in a system highly alien to ourselves, we were unable to apply all aspects of IIT. <direct causation through perturbation> <search of space/time-scale for mechanism> <search for complex> Firstly, IIT has to do with causation. Specifically, in order to test direct causation, IIT requires perturbation of the system into all possible states in order to observe true transition probabilities. Here, we have only used the natural time-evolution of the system, however. Further, IIT also requires search for the ideal spatial and temporal scales for describing the network, such that big-phi is maximised (Oizumi 2014, Hoel 2013 PNAS, Hoel 2016 NoC). These require presumably recording the activity of all neurons in a brain, which clearly is presently infeasible even at the scale of the fly brain.

However, assessing IIT by constructing the IIS and computing Φ in the fly brain provides several advantages. Firstly, using multi-electrode methods provides high-quality population neural signals (both in time and space) unaccessible with any non-invasive measures available in humans. Further unlike bigger brains, small brain of the flies allows us to cover the entire brain. Secondly, given how the computational cost of computing the IIS and Φ grows exponentially as the number of nodes in a system grows linearly, 10^5 of neurons in the fly brain, compared to more complex mammalian brains (mice: 10^8, humans 10^11 neurons) provides a system where computing the IIS across a large majority of neurons in the brain is more feasible to achieve. With regards to fly consciousness, molecular mechanisms of anesthesia, such as decreased action potential amplitudes (Wu 2004 Anesthesiology, Sandstrom 2004 Journal of Physiology) and effects on network dynamics such as reduced feedback connectivity (Lee 2009 consc cogn, Lee 2013 Anesthesiology, Cohen 2018), seemed to be conserved across species. Further fly brains seem to share graph-theoretical characteristics with mammalian brains (Shih 2005 Curr Bio) as well as cellular mechanisms (Littleton 2000 Neuron), and fly LFPs share similarities with human electroencephalographic recordings (Nitz 2002 Curr Bio).

Overall, we have demonstrated applicability of IIT concepts to recordings from the fly brain, a system which is highly dissimilar to our own brains and almost certainly was not considered during the formulation of IIT. Despite this, we show that the predictions of IIT regarding reduced big-phi and a collapsed IIS held up even in such an alien system. Further, though we were limited to very rough brain regions, we demonstrated that II can potentially be used to identify brain regions which may contribute to conscious experience. Our study opens up a new avenue of the theoretical and empirical research on the mathematical basis of conscious phenomenology; unlike previous focus on the system level coarse information and integration, how exactly these integrated information formes an overall shape or structure may inform us more about both level and contents of consciousness.

# Methods

**Experimental procedure**

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. As the data have been published in {Cohen et al., 2018}, here we detail methods directly relevant to the dataset used presently.

Linear silicon probes with 16 electrodes (Neuronexus Technologies) were inserted laterally into the fly’s eye. Probes had an electrode site separation of 25 µm. Recordings were made using a Tucker-Davis Technologies multichannel data acquisition system with a 25 kHz sampling rate. Isoflurane was delivered from an evaporator onto the fly through a connected rubber hose. Actual concentration near the fly body was either 0 vol% (awake condition) or 0.6 vol% (isoflurane condition). Flies in the awake condition responded to air puffs by moving their legs and abdomen, but were rendered inert under the isoflurane condition. Importantly, they regained responsiveness when isoflurane was subsequently removed.

An experiment, as conducted in (Cohen et al., 2018) 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. We used data obtained in 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 downsampled to 1000 Hz. Electrodes were bipolar re-referenced by subtracting neighbouring electrodes, resulting in 15 signals which we refer to as “channels”. The 18 s of data for each condition was split into 2.25 s segments, giving 8 “trials” of 2250 time-samples each. Finally, line noise at 50 Hz was removed using the function 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.

**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. We calculated Φ using PyPhi (version 0.8.1; Mayner, Marshall, & Marchman, 2018, Mayner 2018 PLoS Comp), publicly available from https://github.com/wmayner/pyphi. Below we review mathematical formulation of IIT as it pertains to our work. Complete details can be found in (Oizumi et al., 2014).

To compute the IIS, full and disconnected transition probability matrices (TPMs) are required. To estimate these, we first select a set of N channels of interest. As we used binary time-series, there are 2^N possible states for this set. For each channel X\_n, we computed the probability of being “on” at time *t* + tau given the state of the system at time *t*:

1. p(X\_n(t+tau) = 1 | s(t))

where s(t) represent the state of the system at time t. . In this manner, we computed the TPMs for every possible, 4-channel subset out of the 15 channels. We repeated this procedure for each fly and trial. We selected 4 channels as this gave a reasonable balance between II’s strength of being a multivariate measure and its weakness of exponentially growing computation cost with system size (Mayner 2018).

To compute Φ for a given set of 4 channels, we submitted its associated transition probabilities to PyPhi. Conceptually, PyPhi finds the distance between the probability distribution of transitions specified by the full system with that of the partitioned system (Fig. 5b,d) As there are 2n possible states for set of n-channels, we computed Φ for every state. To obtain a single Φ value for a trial, we averaged Φ across all states, weighting by the number of occurrences of each state (Albantakis, Hintze, Koch, Adami, & Tononi, 2014). The composition of φ values is also provided as output from PyPhi, for both the unpartitioned and partitioned system.

**Classification analysis**

To assess the reliability of the effects of anesthesia, we conducted classification analysis using linear support vector machines (SVM; liblinear). Classification analysis allows us to holistically compare the full, multivariate IIS alongside single small-phi and big-phi values.

We conducted a leave-one-pair out validation procedure for every channel set, where at each validation we removed one wakeful and one anesthetised fly (where the removed flies can be the same fly) and trained the SVM on the remaining 12 wakeful and 12 anesthetised flies. We then tested the SVM on the left-out pair of flies. This was repeated for every possible pairing of wakeful and anesthetised fly to leave out, giving 13\*13 cross-validations. We report the average classification accuracy across all validations.

[standardisation of data for SVM]

To optimise the cost parameter of the SVMs (citation), we repeated the cross-validation procedure at a range of costs, between 2^-20 and 2^20. For all features used for classification, we observed the greatest classification performance at lower cost parameters. Thus, we reported all accuracies obtained from conducting cross-validation at a cost of 2^-20.

**Statistical analyses**

We used linear mixed effects (LME) analysis (Bates, Mächler, Bolker, & Walker, 2015) to test for differences in small-phi and big-phi between wake and anesthesia, and for differences among mechanism orders. LME allows us to account for within-fly correlations among channel sets and avoid averaging across either channel sets or flies. Thus we always included random intercepts for fly and the interaction between fly and channel set as random effects, unless otherwise specified. To test for statistical significance of an effect, we employed likelihood ratio tests, where we compared the log-likelihood of the full model with a model with the effect of interest removed. As the likelihood ratio statistic is chi-squared distributed when two models are equivalent, we report the likelihood ratio statistic with the associated degrees of freedom (chi-squared(d.o.f.)) corresponding to the difference in number of coefficients between the full model with the model with the effect of interest removed, as well as the corresponding p-value. To conduct post-hoc tests, we limited the effect of interest to two levels at a time and report the two-tailed t-statistic and p-value associated with the regression coefficient.

We first employed LME to compare small-phi of the IIS between conscious arousal levels and among mechanism orders. The corresponding model for this test was (in Wilkinson notation; citation):

(2) small-phi ~ condition + order + condition:order + (1|fly) + (1|fly:channel)

Where condition and order are level of conscious arousal (wake or anesthesia) and mechanism order (1, 2, 3, or 4) respectively, and are dummy coded to be treated as categorical variables. The number of observations among mechanism orders differed due to each order having a different number of possible mechanisms (4, 6, 4, and 1, respectively for 1st, 2nd, 3rd, and 4th orders for a set of 4 channels). The term “condition:order” denotes an interaction between level of conscious arousal and mechanism order.

To compare big-phi between wakefulness and anesthesia, we used a similar model:

(3) big-phi ~ condition + (1|fly) + (1|fly:channel)

When comparing classification accuracy across flies across the different feature types (e.g. 1st order mechanism versus 2nd order mechanism), classification accuracy is not nested within fly, thus we only included random intercepts for each channel set:

(4) accuracy ~ feature + (1|set)

Where feature is dummy coded to be one of the 1st, 2nd, 3rd or 4th order mechanisms, the full IIS, or big-phi.

**Acknowledgements**

AL was supported by an Australian Government Research Training Program (RTP) Scholarship. AL and NT were supported by ARC DP180100396. DC and NT were supported by TWCF0199 from Templeton World Charity Foundation, Inc. The opinions expressed in this publication are those of the authors and do not necessarily reflect the views of Templeton World Charity Foundation, Inc. N.T. was funded by the Australian Research Council Future Fellowship FT120100619 and the Discovery Project DP130100194 and DP180104128. B.v.S. was funded by National Health and Medical Research Council Project APP1103923. This work was supported by computational resources provided by the Australian Government through MASSIVE under the National Computational Merit Allocation Scheme.

**References**

Barrett AB, Seth AK. Practical measures of integrated information for time-series data. PLoS Comput Biol 2011; 7. e1001052. Pmid:21283779

Oizumi M, Amari Si, Yanagawa T, Fujii N, Tsuchiya N (2016) Measuring Integrated Information from the Decoding Perspective. PLOS Computational Biology 12(1): e1004654.<https://doi.org/10.1371/journal.pcbi.1004654>

**Supplementary Materials**

~~From the obtained TPM, simple operations of isolating, adding, and dividing rows or columns (Mayner slides) allows us to condition the probabilities given some state, as well as marginalise probabilities across any subset of channels. Thus, we can also condition these probabilities onto not just the state of the full set of channels, but also onto the states of subsets of channels. This allows us to measure the extent to which the state of each subset of channels exists for the full system, by comparing the probability distributions before conditioning onto the state to those after conditioning onto the state.~~

1. Which is called “big phi” (𝚽) in IIT 3.0 [↑](#footnote-ref-0)