Conscious Perception as Integrated Information Patterns in Human Electrocorticography

Andrew M. Haun,1,2 Masafumi Oizumi,2,3 Christopher K. Kovach,4 Hiroto Kawasaki,4 Hiroyuki Oya,5 Matthew A. Howard,4 Ralph Adolphs,5 and Naotsugu Tsuchiya2,6

DOI: http://dx.doi.org/10.1523/ENEURO.0085-17.2017

1Department of Psychiatry, University of Wisconsin-Madison, Madison, WI, 2School of Psychological Sciences, Monash University, Clayton, Australia, 3RIKEN Brain Science Institute, Wako, Japan, 4Department of Neurosurgery, University of Iowa, IA, 5Division of Humanities and Social Sciences, California Institute of Technology, Pasadena, CA, and 6Monash Institute of Cognitive and Clinical Neuroscience, Monash University, Clayton, Australia

Abstract

A significant problem in neuroscience concerns the distinction between neural processing that is correlated with conscious percepts from processing that is not. Here, we tested if a hierarchical structure of causal interactions between neuronal populations correlates with conscious perception. We derived the hierarchical causal structure as a pattern of integrated information, inspired by the integrated information theory (IIT) of consciousness. We computed integrated information patterns from intracranial electrocorticography (ECoG) from six human neurosurgical patients with electrodes implanted over lateral and ventral cortices. During recording, subjects viewed continuous flash suppression (CFS) and backward masking (BM) stimuli intended to dissociate conscious percept from stimulus, and unmasked suprathreshold stimuli. Object-sensitive areas revealed correspondence between conscious percepts and integrated information patterns. We quantified this correspondence using unsupervised classification methods that revealed clustering of visual experiences with integrated information, but not with broader information measures including mutual information and entropy. Our findings point to a significant role of locally integrated information for understanding the neural substrate of conscious object perception.

Key words: consciousness; electrocorticography; face perception; integrated information theory

Significance Statement

What is the link between neural activity and conscious experience? It is clear that experience is generated in the brain, as conscious experience occurs even without sensory inputs, but it is also clear that not everything that occurs in the brain is correlated with consciousness. There must be some phenomenon occurring in brains that is critical for consciousness. In this article, we tackle this issue from a new direction: starting from conscious phenomenology, we derive a novel measure of distributed population neural activity, the integrated information pattern, and find that, when applied to intracranial field potential recordings (electrocorticography, ECoG), this measure can be used to classify the conscious perceptual experiences of human subjects.

Introduction

The contents of conscious experience include our momentary perceptual, emotional, and cognitive experiences, what we are experiencing at this moment. These contents are bound together in a nested, compositional structure (Fig. 1). For example, certain colors and con-
tours are parts of certain surfaces, which are parts of certain objects, which constitute distinct objects, which constitute distinct scenes, which are part of a multimodal conscious percept. Any particular experience is also highly informative in the sense that it takes on a very specific form that excludes an enormous number of other possible experiences. These seem to be fundamental properties of any conscious experience: a highly informative set of many nested parts that are bound into a unified whole.

What could explain how the brain supports the compositional, bound and informative nature of the perceptual aspects of conscious experience? Integrated information theory (IIT; Tononi, 2004, 2012) provides just such a set of explanatory concepts. First, IIT considers the capacity of the brain for state differentiation, which corresponds to the informative nature of conscious experience. The larger the number of distinct states the brain can have, the more informative any particular state must be. Second, IIT quantifies integration of this information across a system by considering how much information is lost if the system is partitioned by disconnecting its parts from one another. Third, IIT claims that different parts of an integrated system will play distinct roles in specifying its total information, and that these roles correspond to parts of conscious experience as a whole. More specifically, IIT proposes that it is the structure of a system's integrated information that corresponds to the exact nature of the brain's conscious contents.

One limitation of the IIT account is that it has been developed in the context of simple model systems, and is computationally intractable for empirically observed neural data (Balduzzi and Tononi, 2008; Barrett and Seth, 2011). An IIT-driven analysis of neural activity thus requires approximated measures of integrated information. One such approximation is derived via the mutual infor-

**Figure 1.** Left, Conscious experience is a multilevel integrated pattern, where, for example, distinct colors constitute shapes, which constitute distinct objects, which constitute distinct scenes, which are part of a multimodal conscious percept. Right, A system of interacting elements (ABCD, lower right corner) generates a multilevel pattern of integrated information. In this example, an integrated system ABCD is supported by AC and BCD, with the latter further supported by BD and CD. IIT proposes that such a pattern of integrated information is isomorphic to the structure of a conscious experience.
information between time-lagged system states (Barrett and Seth, 2011; Oizumi et al., 2016a). Figure 2 illustrates this derivation. The entropy $H(X(t))$ of a multivariate system over a time interval quantifies state uncertainty (the state distribution breadth; Fig. 2A). Mutual information (or predictive information (Bialek et al., 2001) is computed across a short time lag as $I(X(t);X(t-\tau))$: how is the state uncertainty reduced given states at $t-\tau$ in the past (Fig. 2B)? However, mutual information alone does not reflect integration across the system. To evaluate integration we partition the system, estimate $I$ only within the parts, and recombine these estimates. The integrated information $\Phi$ is the divergence of the “true” $I$ and the “partitioned” $I$ (Fig. 2C), where the partition is chosen to minimize the divergence, making $\Phi$ a measure of the irreducibility of the mutual information. Importantly, $\Phi$ is uniquely defined for all parts of a system. For example, the integrated information in a system of three elements, A, B, and C, is exhaustively characterized as the set \{\Phi(AB), \Phi(AC), \Phi(BC), \Phi(ABC)\}; each member of this set identifies some aspect of the system’s integration. We refer to this set of $\Phi$ values as the integrated information pattern or $\Phi$-pattern.

IIT proposes that the structure of a conscious perceptual experience is isomorphic to the integrated information generated in the brain (Fig. 1). If this is true, any change in conscious perception should exactly correspond to change in the pattern of integrated information: if two perceptual experiences are phenomenally similar, they should also be close in the similarity space of integrated information patterns. We tested this hypothesis by measuring $\Phi$-patterns in human intracranial electrocorticography (ECoG) during various distinct visual experiences. The primary manipulation we used in this study is continuous flash suppression (CFS; Tsuchiya and Koch, 2005), which allows dissociation of physical stimulus and conscious perceptual contents. We buttressed our conclusions by extending our analysis to related but distinct paradigms of backward masking (BM) and unmasked stimulus presentation tested in the same subjects. For each stimulus/percept condition we extracted the $\Phi$-pattern from groups of ECoG channels, and gauged the similarity between the patterns and the conscious contents. For comparison, we also extracted mutual information and entropy patterns, which are also compositional information patterns (Fig. 2) but do not reflect integration.
We found that the $\Phi$-patterns mapped onto percepts better than mutual information or entropy patterns. In some patients for whom there were electrodes in object-sensitive areas within the fusiform gyrus, the mapping could be extremely precise and consistent even across multiple stimulus paradigms. We propose that our results support the hypothesized isomorphism between the structure of conscious experience and the integrated information pattern generated by the human brain.

**Materials and Methods**

**Subjects**

We analyzed ECoG recordings obtained in six patients undergoing video EEG monitoring with intracranial electrodes. All patients had “grid” arrays installed over the left ($N = 4$) or right ($N = 2$) lateral temporal lobes, and five also had two or more “strip” arrays installed ventrally on the same side. Patients also had frontal and deep electrodes, which we did not include in our analyses. Recording was not performed within 48 h of major seizures. For the statistical analysis reported in Results, we regarded each row in Table 1 (each combination of subject, experiment paradigm, and lateral/ventral electrode array) as an independent condition. The Institutional Review Board at University of Iowa approved the study (approval number 200112047), and written informed consent was obtained from each patient.

**Psychophysics**

Subjects performed psychophysics experiments during ECoG recording. All subjects were naïve to the tasks, and generally received a block or two of practice trials before data collection began. Completed trial counts for all included subjects are listed in Table 1. More detailed analyses of the tasks (CFS and unmasked tasks) are described elsewhere (Baroni et al., 2017).

The primary task was CFS, where a target face stimulus is presented to one eye while colorful Mondrian (shape noise) images are continuously flashed to the corresponding position in the other eye (Tsuchiya and Koch, 2005). Subjects fixated the stimuli through a custom-made 4-mirror stereoscope. Stimuli were presented on a 19-inch ViewSonic VX922 LCD display (75 Hz refresh rate). Face/Mondrian images subtended $\sim 7.5 \times 10^\circ$ in visual angle. We controlled the experiment and stimuli using Psychtoolbox (Brainard, 1997; Pelli, 1997) with MATLAB (version 7.8) on a PC running Windows XP. In each trial, two temporal intervals were presented (Fig. 3A). Each interval lasted 200 ms, and the two intervals were separated by a random duration between 900 and 1100 ms. In one interval, the target face was presented to one eye; in the other interval, the blank gray field was presented. In both intervals, three distinct patterns of Mondrians were presented (each 66 ms, at 15 Hz) to the other eye. The flashing Mondrian
tended to suppress the target face, especially when the target contrast was lower. As a screening step (to ensure that included subjects understand the task and responded properly), we included only data from experimental sessions whose objective 2AFC hit rate increased with target contrast, and increased with reported visibility. These conditions result in stochastic trial-to-trial visibility of the target face; sometimes the face is consciously seen, sometimes it is not. The subjects’ task was to select which interval contained the face, and to indicate how visible it was on a scale of 1–4. 

**In the BM paradigm,** a target face stimulus, whose emotional expression was either happy, angry, or neutral, was briefly flashed (one frame, ~13 ms) at one of four visual field locations (upper-left, upper-right, lower-left, lower-right), which was immediately replaced by a gray blank screen. The same display settings as the CFS experiment, except for the mirrors were used. The subject’s task was to identify the location of the face target among four possible locations, and also to identify its emotional expression among three possible labels (happy, fearful, and neutral). 

**Figure 3.** Paradigms for stimulus presentation. A, CFS task. Each trial consisted of a temporal sequence of two stimulus intervals, separated by a random interstimulus interval (ISI, 900–1100 ms). In one eye, each interval contained three flashes of colorful Mondrian patterns; in the other eye, one interval contained a face image of variable contrast. These conditions result in stochastic trial-to-trial visibility of the target face; sometimes the face is consciously seen, sometimes it is not. The subjects’ task was to select which interval contained the face, and to indicate how visible it was on a scale of 1–4. 

B, BM task. After a random fixation delay, subjects saw an array of four noise patches, one of which contained a face image (the upper left, in the illustration), for 13 ms. After a variable SOA, another array of noise patches was presented to reduce the visibility of the face. The subject’s task was to identify the location of the face target among four possible locations, and also to identify its emotional expression among three possible labels (happy, fearful, and neutral). 

C, Unmasked conditions included a one-back memory task, in which subjects paid attention to the category of the stimuli, or a simple fixation task, in which they detected a change of color orientation of the fixation cross. In both tasks, faces and other objects were presented for 500 ms without any masks, with trials separated by a blank interval (500 ms for the fixation task, 1000 ms for the one-back task). 

D, From subjects’ performance on a task (correct/incorrect, ratings of visibility, identification of expression), we can reasonably infer their percept on each trial of an experiment. We used subjects’ performance to divide trials into the percept categories shown here: faces (and inverted faces), houses, tools, Mondrians, and noise.
of the target face with two button presses. In absence of an explicit visibility judgment, we coded trials where subjects responded correctly for both location and emotion as visible, which can happen only once in twelve random guesses. We also treated trials where subjects incorrectly identified the location as invisible. Trials where the subject correctly identified the location but not the emotion were excluded from analysis for this study. We did not include the face emotion as an independent variable in any analyses for this study. Summary measures of BM performance are shown in Figure 4 (Right).

In the unmasked paradigms (UNM), stimuli were presented on a continuous cycle while subjects fixated the center of the display. Stimuli included upright faces, upside-down faces, houses, line drawings of tools, and Mondrian patterns used in CFS (but not flickering). Subjects indicated either the change of the fixation cross color (“fixation task”) or, in separate experimental blocks, the repeat of stimulus category from trial to trial (“one-back task”). Since most trials did not require a response, UNM blocks were included for analysis if the subject made responses during the task, but no accuracy threshold was imposed. For the analyses in this article, we did not distinguish between the fixation and one-back tasks.

Of these three tasks, only the CFS included a metacognitive judgment. Within a subset of CFS trials, subjects received physically identical visual input; yet they consciously experienced seeing the face target only on some of the trials. By dissociating stimulus parameters from conscious perception, this type of paradigm has been successfully used to isolate the neural correlates of consciousness (NCC; Koch, 2004). We include the other tasks as complementary to the CFS task to test the generality of our claims across different tasks. With regard to our classification of BM trials: nonconscious emotion discrimination has been reported in blindsight ( Pegna et al., 2005) and in healthy subjects (Jolij and Lamme, 2005). However, we required correct discrimination of both location and emotion of the face targets, reducing the likelihood of including trials with a perceptually invisible face. In the UNM task, we presented the stimuli for a relatively long duration (500 ms) at fixation with no masks, no misdirection of expectation, and no attention demanding tasks. In such a situation, target perception is extremely unlikely to be nonconscious.

**Computing \( \Phi^* \)**

\( \Phi^* \) is a measure of integrated information in a candidate subsystem \( X \), derived in the following sequence of computations (I–III). For detailed mathematical derivation of \( \Phi^* \) (Oizumi et al., 2016a). A Matlab toolbox for computing \( \Phi^* \) may be found at https://figshare.com/articles/phi_toolbox_zip/3203326 (RRID:SCR_015549).

1. The state of an \( n \)-channel subsystem at time \( t \), which we denote as \( X(t) \), is an \( n \)-dimensional vector, where its \( i^{th} \) dimension specifies the bipolar re-referenced voltage for the \( i^{th} \) channel. To estimate the uncertainty about the states of the mechanism, we employ the concept of entropy, under an assumption of Gaussian distribution of these states (Barrett and Seth, 2011; Oizumi et al., 2016a):

\[
H(X) = \frac{1}{2} \log (\Sigma(X)) + \frac{1}{2} \log (2 \pi e),
\]

where \( \Sigma(X) \) is the covariance matrix of \( X \), estimated over the time interval \([t-T/2, t + T/2]\). \( T = 200 \) ms throughout the article; we found 200 ms to be a realistic range of the temporal window for conscious perception and a good compromise between better temporal specificity and better estimation of the covariances. The \( (i,j)^{th} \) element of \( \Sigma(X) \) is the covariance between channel \( i \) and \( j \) of \( X \) over the time interval. \( \Sigma(X) \) is the determinant of \( \Sigma(X) \), known as the generalized variance (Barrett et al., 2010), as it describes the \( n \)-dimensional spread of instantaneous states of \( X \). The more different states \( X \) takes over the time interval, the more uncertain we are about its state at any time \( t \) within the interval.
II. Next, we estimate reduction in uncertainty about the mechanism’s states at \( t = X_t \) given its past states \( \{X_{t-	au}, \ldots, X_{t-t} \} \) using the concept of mutual information \( I \):

\[
I(X_t; X_{t-	au}) = H(X_t) – H(X_t|X_{t-	au}) \text{ ,}
\]

where \( H(X|X_{t-	au}) \) is the conditional entropy of the mechanism \( X \) across the delay \( \tau \). Under the Gaussian assumption, the conditional entropy is given by

\[
H(X_t|X_{t-	au}) = \frac{1}{2} \log \left( (\sum(X_t|X_{t-	au}) \right) + \frac{1}{2} n \log (2\pi e). \tag{3}
\]

The covariance matrix of the conditional distribution, \( \Sigma(X_t|X_{t-	au}) \), is

\[
\Sigma(X_t|X_{t-	au}) = \Sigma(X_t) – \Sigma(X_t, X_{t-	au})\Sigma(X_{t-	au})^{-1}\Sigma(X_t, X_{t-	au})^T, \tag{4}
\]

where \( \Sigma(X_t, X_{t-	au}) \) is the cross covariance matrix between \( X_t \) and \( X_{t-	au} \), whose element \( \Sigma(X_t, X_{t-	au}) \) is given by covariance between \( i \)-th element of \( X_t \) and \( j \)-th element of \( X_{t-	au} \).

The way we use mutual information here is similar to predictive information or auto-mutual information (Brenner et al., 2000; Gómez et al., 2007; Julitta et al., 2011). I \((X_t; X_{t-	au})\) is a measure of the information that current states have about their own past states.

III. Integrated information \( \Phi^* \) over the subsystem \( X \) is information that cannot be partitioned into independent parts of \( X \) (For simplicity, we remove \( t \) and \( t-\tau \) from \( X \) for the explanation of \( \Phi^* \) here). To identify integrated information in a subsystem, we estimate the portion of the mutual information that can be isolated in parts of \( X \). The process consists of first defining the parts of \( X \) (a "partition") and then estimating the total mutual information under the assumption that the parts of a subsystem are independent. An estimate of “disconnected” \( I' \) is called mismatched information and denoted as \( I' \) (Oizumi et al., 2016a; Oizumi et al., 2010).

We compute \( I' \) for every possible partition of \( X \). As an example, if \( X \) is made up of four ECoG channels, there are 14 possible partitions of the subsystem (e.g., \{abcd\}, \{ab|cd\}, \{abc|d\}, etc.), excluding the “trivial partition” where all \( n \) elements are together in a single part (e.g., \{abcd\}). For each partition, we obtain \( \Phi^* = I - I' \). We select the partition \( P \) that minimizes the normalized \( \Phi^* \) value, as defined previously (Balduzzi and Tononi, 2008):

\[
N_P = (m-1) \times \min_k \{ H(M^k) \} \tag{5}
\]

\[
MIP = \arg\min_{\Phi^*} \left\{ \frac{\Phi^*_i}{N_P} \right\} \tag{6}
\]

where \( m \) is the number of partitions and \( M^k \) is the \( k \)-th part of the system \( X \). The normalization term \( N_P \) counterbalances inevitable asymmetries introduced by computing \( \Phi^* \) across variable numbers of partitions of unequal size (Balduzzi and Tononi, 2008). The partition that minimizes normalized \( \Phi^* \) is called the minimum information partition (MIP). The MIP reveals the weakest link between the parts of \( X \). The integrated information of the subsystem is defined across the MIP as \( \Phi^*_{\text{MIP}} = I - \Phi^*_{\text{MIP}} \) (throughout this article, \( \Phi^* \) refers to \( \Phi^*_{\text{MIP}} \)).

For stable estimation of covariance and cross-covariance, we used a shrinkage approach (Schäfer and Strimmer, 2005). By computing covariance and cross-covariance matrices separately for each trial and averaging these before computing the entropy, we estimated \( \Phi^* \) for bins of trial data. For the classification analyses, bins consisted of three trials each (randomly selected from a given percept/stimulus category); for the exemplar \( \Phi^* \)-pattern illustrations shown in the Results section, all trials belonging to a particular percept/stimulus category were used to produce averaged covariance and cross-covariance matrices.

The \( \Phi^* \)-pattern

To mirror the nested, multiorder structure of perceptual experience, we measured \( \Phi^* \) for every subsystem within a selected system of ECoG channels. For a system of four channels (ABCD), there are 11 such subsystems: ABCD, ABC, ABD, ACD, BCD, AB, AC, AD, BC, BD, and CD. The resulting set of \( \Phi^* \) values describes a pattern of overlapping subsystems that may or may not integrate information within the system. We refer to this set of integrated information as the \( \Phi^* \)-pattern. A structure based on the same composition of subsystems can be obtained using mutual information (I) or entropy (H). I and H sum linearly as subsystems are combined unlike integrated information, which may decrease as the size of the subsystem increases. I/H patterns were used as comparison for \( \Phi^* \)-pattern in statistical analyses (in Results, section ‘Classifying Conscious Contents’).

Channel set selection procedure

We reasoned that where a stimulus evokes information integration, the pattern of the integration should identify the content of a conscious percept, especially in visual areas. To test this hypothesis, we conducted the following procedure for each specific experimental setting (a particular stimulus task, as viewed through a particular brain region of electrode implantation, either ventral or lateral regions, in a particular subject). The “max evoked \( \Phi^* \)” set of channels \( X \) is the subsystem and time lag \( \tau \) where the largest average \( \Phi^* \) was evoked (averaged over all three trials for all stimulus/percept categories):

\[
\{\hat{X}, \tau\} = \arg\min_{X, \tau} (\Phi^*(X, \tau)), \tag{7}
\]

where \( \hat{\Phi}^*_i \) was the largest average evoked \( \Phi^* \) for the \( i \)-th subsystem \( (i = 1, 2, \ldots, 11) \) within a channel set \( X \) for a particular \( \tau \):

\[
\hat{\Phi}^*_i (X, \tau) = \max_i (\Phi^*_i (X, \tau)), \tag{8}
\]

and the mean evoked \( \overline{\Phi}^*_i \) was a subset of the mean post-stimulus and mean pre-stimulus \( \Phi^* \) for the \( i \)-th subsystem for a particular \( \tau \):

\[
\overline{\Phi}^*_i (X, \tau) = \frac{1}{300} \sum_{t=-300}^{300} \Phi^*_i (t) + \frac{1}{100} \sum_{t=-100}^{100} \Phi^*_i (t). \tag{9}
\]
In Equation 9, \( t \) is the center of a 200-ms interval of ECoG data, and the \( \Phi^n \) estimates are averaged over all three-trial bins for all stimulus/percept categories. We evaluated \( \Phi^n \)-patterns at each searchlight, and for each of four \( \tau \) values (1.5, 3, 6, and 12 ms). As an ROI for further analysis, we chose one channel set that contained the subsystem with the highest \( \Phi^n \) regardless of stimulus/percept (see Materials and Methods for details). The selected regions for classification analysis (Figs. 9, 11) are marked in red. Estimated anatomic region of the max evoked-\( \Phi^n \) channel set are given in Table 1. Marker colors encode the magnitude of the maximal evoked \( \Phi^n \) at the centroid of each system (values above 0.1 are all given the color yellow).

The max evoked \( \Phi^n \) set for the CFS condition was next to the system depicted in Figure 5.

**Unsupervised clustering for assessing the similarity of informational patterns**

We used unsupervised classification for all available subjects within each stimulus paradigm and within recording locations (either lateral or ventral temporal surface), comparing classification performance across the three types of information pattern (\( \Phi^n \), I, and H). As input features to the clustering, we reduced the dimensionality of the input features into the first 4 multidimensional scaling coordinates derived from (Pearson) correlation-distance matrices (Kriegeskorte et al., 2008), and used the nearest-neighbor algorithm for percept classification.

Measurement of conscious percept classification began with defining categories for each experimental stimulus interval. For CFS, we performed two analyses. In Figure 11C, we used only the intervals that contained mid-contrast target faces (so that visual input was the same across all trials) then categorized each trial according to the reported percept in each trial as either visible face (visibility 3 or 4) or visible Mondrian (visibility 1 or 2; Fig. 3). For Figure 11B, we used all available intervals for CFS; the percept categories were either visible face (face-present intervals with visibility 3 or 4 regardless of the
contrast of the face) or visible Mondrian (face-present intervals with visibility 1 or 2 as well as face-absent intervals). For BM, they were either visible face (trials with correct discrimination in both location and emotion regardless of SOAs, where random responses would result in only 8.3% accuracy) and Visible Noise (trials with incorrect location discrimination regardless of SOAs as well as catch trials where no face was presented). For UNM stimuli/percepts were either faces (including upright and inverted faces) or nonfaces.

We assessed the clustering with a cross-validation procedure. Cross validation removed any possible bias due to unequal number of trials due to percept. First, using 70% of randomly sampled trials within each percept category as a training set, we determined the center of gravity for each category. Second, using the remaining 30% of trials as a test set, we obtained a series of hit and false alarm rates by gradually changing a criterion distance from the category center. We varied the criterion radius for each category from the mean estimated from the training set, counting the number of same-category bins within the radius as “hit” and other-category bins as “false alarm.” By smoothly varying the radius from 0 to infinity, the proportion of both Hit and False Alarm changes from 0–1, yielding a receiver-operating characteristic (ROC) curve for each class. We averaged the area under the ROC curve (AUC) over all categories as the measure of the classification accuracy. The procedure was repeated 20 times, with random resampling of train/test bins. For statistical analysis, each AUC value was converted into a zAUC score through the inverse cumulative normal distribution, which removes the 0–1 bounds of the AUC value and is more appropriate to ANOVA.

Results
The structure of integrated information
We measured information patterns (I, H, and I) at every location of a four-channel searchlight, in each of six subjects, over 22 conditions (different tasks and electrode
settings: Fig. 5) for a range of \( \tau \) (time lag) values from 1.5–12 ms. All computations were performed over 200-ms time windows spaced 100 ms apart over a period from 400 ms before stimulus onset to 1000 ms after onset. To select regions of interest, we focused on maxima in evoked \( \Phi^* \), i.e., the searchlight locations with maximal increase in \( \Phi^* \) after stimulus onset, regardless of the stimulus category.

We first present a detailed look of an integrated information pattern in a single subject (S153) for the CFS task, in a group of channels located over the right fusiform gyrus, a region that has a known association with conscious perception of faces (Tong et al., 1998; Puce, Allison, and McCarthy, 1999; Parviz et al., 2012; Rangarajan et al., 2014). The process is illustrated in Figure 6 for a system of four channels (Fig. 6A). These four channels (with \( \tau = 3 \) ms) contained the highest evoked \( \Phi^* \) of any searchlight location in this subject (Fig. 5, S153). Figure 6B plots the average time course of raw bipolar referenced voltages (no baseline correction) during high-contrast face intervals in the CFS task, where subjects felt faces were highly visible (visibility ratings of 4). We computed \( \Phi^* \) for every subset of at least two channels from this system: for a system of 4 channels this yields 11 subsystems: six two-channel subsystems, four three-channel subsystems, and a single four-channel subsystem. Figure 6C–E shows the time courses of the quantities that underlie \( \Phi^* \), computed for each of 11 subsystems. The most general quantity is the entropy \( H \) (Fig. 6C). From the entropy \( H \), we subtract the conditional entropy, yielding the mutual information \( I \) (Fig. 6D). The integrated information is the difference between the mutual information and the sum of the mutual information of its parts: \( \Phi^* = I - I' \) (evaluated at the MIP of the subsystem; Fig. 6E).

Most subsystems show a similar time course for \( H \) and \( I \) (in the example system of Figure 6 as well as other observed systems), with entropy and mutual information decreasing a few hundred milliseconds after stimulus onset. In contrast, \( \Phi^* \) has an idiosyncratic time course, strongly dependent on the specific subsystem. For most subsystems, \( \Phi^* \) remains near zero regardless of the visual stimulus; for others, it may start high and drop after a particular stimulus is presented (Fig. 6E, subsystem BCD); and for other subsystems, \( \Phi^* \) starts low and increases after stimulus presentation (Fig. 6E, subsystems BD and ACD).

**\( \Phi^* \)-patterns as graphs**

The \( \Phi^* \)-pattern is not just a vector of integrated information values; each value has a necessary relation to some of the others, in the form of an inclusion or intersection relation. This property of the \( \Phi^* \)-pattern distinguishes it from more orthodox measures that aim to reduce neural activity to independent components: the components of the \( \Phi^* \)-pattern are by design not independent. The inclusion relations in a \( \Phi^* \)-pattern can be illustrated in the manner of a Hasse graph (Fig. 7A). In the graph, we place the highest four-channel subsystem ABCD at the origin (the red colored node), surrounded by the three-channel subsystems in black, which are further surrounded by the two-channel subsystems in blue. The edges represent adding or subtracting one channel to a subsystem. For example, three-channel subsystem ACD is connected to subsystems AC, CD, AD via edges as well as ABCD. The pattern is more easily appreciated in a 3D rendering (Video 1).

The four-channel \( \Phi^* \)-pattern in Figure 7 illustrates an important property of \( \Phi^* \) that distinguishes it from \( H \) and \( I \): nonlinearity and nonmonotonicity with respect to the size of the subsystem. Here the three-channel subsystem ACD attains the highest \( \Phi^* \) of all 11 possible subsystems. The graph illustrates how subtracting from or adding to a subsystem can reduce the integrated information. The \( \Phi^* \) of ABCD is less than that of ACD because there is a relatively weak interaction between B and ACD. In this case, the MIP for ABCD is between B and ACD.

Figure 8 shows \( \Phi^* \)-patterns obtained in three separate experiments: CFS, BM, and the unmasked one-back task (For comparisons of these patterns, see Video 2 to view \( \Phi^* \)-patterns in a 3D perspective). \( \Phi^* \)-patterns in the first
two columns are constructed from trials where the subject clearly perceived a face (94 trials with high visibility ratings in CFS, with hit/miss counts identified in the inset text, 68 trials with correct in both location and emotion discriminations in BM, and 81 trials of unmasked upright or inverted faces in the one-back task), whereas the last two columns are not (95 low visibility rating CFS trials, 28 incorrect location discrimination BM trials, and 38 Mon-
CFS trials (Face Vis:4, Contrast:3)

Video 1. Four-channel \( \Phi^- \)-pattern (from Figs. 6, 7), rotated through 3D to clearly illustrate its construction. The \( x-y \) coordinates are arranged to illustrate the layout of the Hasse graph that connects all the subsystems. The \( z \)-coordinate is the magnitude of \( \Phi^- \) for each subsystem. This pattern is generated by a system of four channels over subject 153’s posterior fusiform gyrus. [View online]

Figures 6-7 were derived by combining all available trials for each stimulus/percept condition; if we break the data down and derive a \( \Phi^- \)-pattern for just a few trials’ worth of data, the mapping to stimulus/percept still holds: Each dendrogram leaf in Figure 9 corresponds to a \( \Phi^- \)-pattern for a bin of three stimulus/percept-matched trials. \( \Phi^- \)-patterns corresponding to visible face percepts cluster together and separately from \( \Phi^- \)-patterns generated during other percepts. For comparison, we conducted the same analysis for patterns of mutual information and entropy (I and H; middle and lower panels, respectively) across the same set of subsystems as in the \( \Phi^- \)-pattern; these appear to yield poorer clustering. This suggests that refining this particular neural system’s activity to a pattern of integrated information results in a closer structural match to conscious perception of faces across different task contexts, and reasonable classification of other percept categories as well. However, these “by eye” analyses of the mapping between \( \Phi^- \)-patterns and conscious contents are merely suggestive; in the next section, we apply objective clustering analyses over a group of subjects.

Classifying conscious contents with integrated information

To evaluate the mapping between percept and \( \Phi^- \)-pattern, we constructed matrices representing dissimilarity (the Pearson distance) of \( \Phi^- \)-patterns derived from bins of three stimulus/percept-matched trials. As a dimension-matched comparison, we also measured dissimilarity for entropy (I) and mutual information (H) patterns. We used these dissimilarities as the basis of a clustering analysis to determine how well the different types of information pattern mapped onto the known percept/stimulus condition for each bin of trials (see Materials and Methods for details). Importantly, there was no training step to optimize weights for this unsupervised classification: the pattern coordinates served directly as unweighted feature vectors. This is important because our motivation for this representation similarity analysis was not to find the most accurate decoder. Rather, we wanted to test if a hierarchy of integrated information might show direct isomorphic correspondence with conscious percepts (see Discussion).

We obtained local \( \Phi^- \), I, and H patterns at hundreds of “searchlight” systems in each subjects’ electrode arrays, with each pattern computed on a set of three percept/stimulus-matched trials’ worth of data (see Materials and Methods for details). For each experiment condition (subject/electrode array/task) we conducted the classification analysis separately at every searchlight location. Figure 10 summarizes the searchlight classification results, combining AUC scores over all experiment conditions by expressing the likelihood of each pattern type given a specified AUC range. In other words, if a searchlight system yielded a certain AUC, what is the probability that it was with an integrated information pattern, a mutual information pattern, or an entropy pattern? A high degree of pattern similarity to conscious percepts is very rare (most searchlight locations yield chance performance), but...
Figure 9. Pattern clustering for the channels featured in Figures 6-8 in subject 153, across 364 bins of three trials each. Three dendrograms are illustrated based on the patterns of $\Phi^*$ (top), mutual information $I$ (middle), and entropy $H$ (bottom). The dendrograms represent relative (Euclidean) distance between patterns. Four coarsely defined percept categories are conscious perception of faces (blue), masking stimuli (red), houses (green), and “tools” (yellow). The red category, conscious perception of masking stimuli, corresponds to the two right columns in Figure 8, and it consists of UNM Mondrian trials, BM invisible trials (location discrimination incorrect) and catch trials (no face presented), CFS invisible face-present intervals (visibility 1 or 2), as well as CFS face-absent intervals (Fig. 3). Thus, we analyzed all available trials in BM regardless of SOAs and CFS regardless of presence or absence of the face as well as all face contrasts. Below each dendrogram are two bands of colors: the upper band identifies the percept category for each dendrogram leaf (trial bin), and the lower band identifies the stimulus paradigm: CFS (dark gray), BM (light blue), and UNM (light magenta).

Figure 10. The likelihood of each pattern type given specific AUC ranges, over all searchlight locations, all stimulus paradigms, and all subjects. AUC is taken as the average between 200 and 400 ms after stimulus onset. Likelihood is the incidence of a pattern type at the given AUC, divided by the total number of sets yielding the given AUC (indicated by the $N$ over each group of bars). Thus, the chance level is 0.33 (i.e., one of three pattern types), indicated by the horizontal gray line. Each column is subdivided according to the contribution of six subjects, the numbers over each column indicate the number of subjects contributing to each pattern type at that AUC range. This analysis does not rely on a particular channel selection strategy, and emphasizes that accurate matches with percepts (AUC $> 0.6$) are predominantly found with the $\Phi^*$ pattern.
but when similarity is high Figure 10 shows that it is most likely to be with a $\Phi^*$-pattern.

Next, we selected the searchlight system that generated the largest evoked $\Phi^*$ regardless of stimulus/percept for each distinct experiment condition (as listed in Table 1). Note that using the same criterion, we selected the system featured in Figures 6-9 for S153; see Materials and Methods for details). Figure 11A shows the average AUC time course for these systems based on integrated information ($\Phi^*$), mutual information ($I$) and entropy ($H$); average classification is best with $\Phi^*$-patterns. Figure 11B shows the same result broken down by stimulus paradigm: integrated information outperforms in both masked and unmasked stimulus paradigms.

Importantly, the same result is obtained when classification is performed only on stimulus-identical CFS face-present intervals (Fig. 11C). During these intervals, the physical input to the brain was identical: mid-contrast faces paired with Mondrian masks, and the only variable was the subjective visibility of the face; the best classification of subjective visibility was obtained with the $\Phi^*$-patterns, in each of three observers (a fourth CFS observer, S154, did not supply enough data to compute AUC for any single contrast level).

We used the poststimulus-onset zAUC scores (see Materials and Methods) over all conditions (combinations of six subjects, three tasks, and two types of electrode installation; see Materials and Methods), with pattern type, subject ID, and task as fixed factors and poststimulus time point as a covariate for an ANOVA. There was a main effect of pattern type ($F_{(2425)} = 15.332, p < 0.001$), with significant (Bonferroni-corrected) post-hoc pairwise differences between zAUC derived from $\Phi^*$ versus I patterns ($p < 0.001$), and $\Phi^*$ versus H patterns ($p < 0.001$), but not between I and H patterns (these comparisons reflect differences between the poststimulus time courses in Fig. 11A). There were also significant main effects of subject ID ($F_{(6425)} = 47.802, p < 0.001$) and task ($F_{(6425)} = 14.866, p < 0.001$). Except for the pattern-type-task interaction, all other factorial interactions were significant, likely reflecting the large heterogeneity of the data conditions.

In the preceding analysis, we simply reasoned that where information is integrated (or causal interactions are evoked) in response to a stimulus, its pattern should map closely to the reported experience. Although Figure 11 shows the advantage of $\Phi^*$ over I or H in most systems in the ventral and lateral cortex, using the above procedure...
to select ROIs for statistical analysis may give AUC(\(\Phi^*\)) an unfair advantage over AUC(I) and AUC(H). In fact, AUC(\(\Phi^*\)) is correlated with evoked \(\Phi^*\). To evaluate this problem, we also selected channel sets based on the minimum or maximum evoked I and H. Regardless of the selection criteria, however, the highest AUC was obtained with \(\Phi^*\)-patterns (Fig. 12), although the overall level of classification tends to be poorer than when the criterion is based directly on \(\Phi^*\).

**Discussion**

We have presented a novel analysis which allowed us to evaluate the degree of similarity between integrated information patterns in neural activity and specific conscious experiences. Some theories of consciousness, most prominently the IIT (Tononi, 2004), predict that conscious experiences should closely match with a hierarchical pattern of causal neural interactions. Although our investigation of this hypothesis cannot be posed as a test of the IIT per se (see below), we did find that the integrated information pattern in human cortex closely matched with psychophysically-established perceptual contents. This finding generalized across stimulus paradigms and included conditions where physical stimuli were dissociated from perceptual experiences. Furthermore, \(\Phi^*\)-patterns map more closely to perceptual states than do entropy and mutual information patterns, despite their overlapping derivation.

The advantage of \(\Phi^*\) over entropy and mutual information is likely due to \(\Phi^*\)’s isolation of integration. Entropy quantifies only the instantaneous uncertainty of the neural states, understandable as equal-time interactions (Oizumi et al., 2016b; Fig. 2A). Mutual information quantifies causal (time-lagged) interactions in these distributions, i.e., how past states affect future states, including both within and between channels (Fig. 2B). \(\Phi^*\) also quantifies time-lagged interactions but only those that tie groups of channels together (Fig. 2C). Thus, our results can be summarized as showing that neither the pattern of equal-time interactions (H) nor of time-lagged self-interactions (I) resemble patterns of conscious percepts as well as causal cross-interactions (\(\Phi^*\)). This suggests a critical role of causal interactions across the system of neuronal activity for understanding how conscious phenomenology corresponds to neural systems. Indeed, when we applied our analyses to patterns of self-interactions (i.e., patterns of single-channel based entropy and mutual information; Fig. 13), the pattern did not correspond to conscious percepts at all.

Another way to understand the advantage of \(\Phi^*\) is as noise reduction: each step in the derivation of \(\Phi^*\) involves not the construction of new information from the neural data, but “deletion” of noise, thus improving the performance of the clustering algorithm. When mutual information is computed from entropy, it is by subtracting the part of the system noise that is only within parts of the system. Derivation of \(\Phi^*\) can then be seen as reduction of the system noise to only those portions that persist across the time lag, and across the parts of the system. This residual maps closely onto the quality of a conscious percept.

In the notable case of subject 153 who provided a large number of trials in the CFS and other experiments, we observed a similar \(\Phi^*\)-pattern generated in different stimulus paradigms when the subject experienced a “face” percept. The elements generating the \(\Phi^*\)-pattern, channels of neural activity in the right fusiform gyrus, have long been associated with face percepts on the basis of patterns of responsivity and perceptual consequence of ele-
Figure 13. To isolate the contribution of self-interactions, we repeated the main analysis described in Figure 11A, but we used only H and I derived from each channel (i.e., {H(A), H(B), H(C), H(D)}, etc.). Using these 4D patterns, instead of 11D pattern, we performed the same analysis, but the results were very poor, consistent with a critical role of cross-channel causal interactions.

Interpretation of integrated information patterns

In theory, the low-order $\Phi^*$ values for local parts of the system would correspond to particular features or dimensions of a larger experience; the higher-order $\Phi^*$ values would then correspond to the binding of these features. For example, suppose that connected but distinct neural populations encode the presence of particular facial features (an enormous simplification; Freiwald et al., 2009; Henriksson et al., 2015; Chang and Tsao, 2017), with one population encoding eyes (population E), another encoding nose (population N), another encoding mouth (population M), and another encoding a group of features (population G). When a face is seen, each population will be active, but the features must be somehow bound together in a whole to correspond to the holistic experience of a face composed of features; that is, there must be a physical substrate that reflects the integration of facial features. For such a system of neural populations, when a face is seen, activity in mouth might be moderately integrated with activity in nose (i.e., $\Phi^*[M,N]>0$), nose with eyes ($\Phi^*[N,E]>0$), all three features together ($\Phi^*[M,N,E]>0$), and any of these should be strongly integrated with the group population ($\Phi^*[E,G]>0$, $\Phi^*[M,G]>0$, etc.); meanwhile, mouth and eyes populations may not be integrated at all (i.e., $\Phi^*[M,E] \approx 0$). The $\Phi^*$-pattern generated by these populations during visual experience of a face would therefore have a particular structure isomorphic to the complex experience. While we have no way of ascribing particular functional roles to the ECoG channels in our dataset, this scenario suggests a plausible interpretation for $\Phi^*$-patterns.
The idea of an informational hierarchy having a close relation to the substrate of consciousness may seem familiar: computational theories (e.g., Marr, 1982) and more recently predictive coding theories (e.g., Hohwy, 2013) of the mind heavily rely on the notion of hierarchical substrates. In these schemes, a “high-level” neuron may receive inputs from “low-level” neurons, and feed back onto the lower level to adjust the sensitivity of its inputs. However, the neurons themselves cannot know their location within the hierarchy. Likewise, they cannot know which inputs are from high- or low-level neurons. The hierarchy is only there in the description of the system by the experimenters. IIT does not take on this kind of extrinsic viewpoint: the hierarchical structure of a system emerges intrinsically as the integration of the system’s parts. Thus, we did not assume any hierarchy of ECoG channels, but instead extracted a hierarchical pattern intrinsic to the activity across those channels. This is an important property afforded by IIT that does not seem to be obtained from hierarchical models of information processing in the brain.

Considered in this way, even detached from IIT, the notion of an intrinsic hierarchy of integrated neural structures may prove to be extremely fruitful for future neuroscience research. In fact, current perspectives on neural processing do seem to be moving toward high-dimensional descriptions of how connected brain systems accomplish the jobs of consciousness, perception, and cognition. Looking at synapse-level connectivity of large groups of single neurons (Reimann et al., 2017) have recently shown how neural systems can be understood in terms of their topological features, treating them as sets of overlapping groups of connected neurons. In another study, Lord et al. (2017) have described how a different set of methods to distinguish “integration” (the holistic property of a neural system) from “segregation” (the way the parts are differentiated from one another) can be used to understand psychiatric disorder and other aspects of whole-brain function.

A novel approach toward finding the neuronal isomorphism of consciousness

As we emphasized in Results, we intentionally avoided “optimizing the classifier” by finding the best weights to achieve higher AUC. While such decoding analyses have their own advantages in the search for NCCs (as in the same data set presented in this paper; cf. Baroni et al., 2017), there is an important caveat in such an approach: the assumption of the decoder. Even if one can perfectly identify conscious perception based on patterns of neuronal activity, the interpretation of a vector of regression weights, for example, is entirely dependent on the presence of a decoder, usually a tool in the hands of the experimenter. What we learn from this kind of analysis is that the information necessary for consciousness is available in (e.g.) a certain brain area. But there is no clue as to what it should feel like, why should one vector of weights feel so different from another? A hierarchical structure of integrated information does not pose the same problem, because it is hypothesized a priori as the structure of experience, serving as a theoretical basis for the experiment.

Accordingly, our goal in this study was not to find the optimal classifier for these data; our goal was to test the prediction that an integrated information pattern in the brain corresponds to the contents of consciousness. Our approach is therefore distinct from the traditional “search for NCCs.” We aimed not to build a perfect classifier that presupposes an ideal homunculus who reads the encoded information, but to test if a hierarchy of causally connected and integrated information can possibly relate to contents of consciousness. Our results provide an affirmative answer.

Summary

A conscious experience is intrinsic, existing for itself, not for some outside observer, $\Phi$ measures how a system specifies its own state, not an external state; a conscious experience is integrated, irreducible, more than the sum of its parts, $\Phi$ measures how the system as a whole specifies its own state, above and beyond its elements; and a conscious experience is compositional, just like the hierarchical $\Phi$-pattern. Based on these theoretical parallels, there is strong a priori reason to expect $\Phi$-patterns should closely correlate with perceptual experience. This type of theoretical background is lacking in many investigations of the NCC. In this regard, our study can be seen as a test of a theory, albeit limited and indirect as we acknowledge above, rather than the traditional search for NCCs.

In this study, we have tried to keep things simple, concentrating on four-channel systems, and have been limited by the (experimentally) incidental placement of recording sites. Our results should be seen as a first step in the direction of a new empirical research framework in consciousness science, where we assess the degree of structural similarity between conscious phenomenology and theoretical construct, with empirical data as a bridge between the two (Tsuchiya et al., 2016; Tsuchiya, 2017). Future studies may use data that is obtained explicitly for the purpose of extracting integrated information patterns, over larger and larger data windows. If there is indeed an isomorphism, which the current study can only suggest, then the structure of a conscious experience should be similar to the structure of the theoretical construct, pointing to a need for assessment of phenomenological structure alongside collection of neural evidence.

Looking further ahead, IIT would require that for a local $\Phi$ pattern to be a part of conscious experience as a whole, it must be a part of a much larger pattern extending across multiple cortical regions. Our procedure did not measure this superstructure, called a complex in IIT (Tononi et al., 2016), another reason that our procedure cannot be taken as a direct test of IIT. We see several other immediate avenues of development that should follow from our study; first, we would propose that new tests of the $\Phi$-pattern concept should include conditions of constant stimulation under varying levels of consciousness (e.g., comparison of $\Phi$-patterns during wakefulness and anesthesia), similar types of $\Phi$-pattern should be observed during dream states as during wakefulness, but different types (with much lower $\Phi$ magnitude) should be observed during dreamless sleep or deep anesthesia.

With the methodological framework we provided here, it will be possible in the future to test a subset of the theory’s predictions, that hierarchies of integrated infor-
mation are isomorphic to the structure of conscious contents, i.e., that the binding problem may be resolved by this approach. The 4p-pattern concept can be applied to other data types in other experimental manipulations combined with neural recording and stimulation, under various conscious states (e.g., awake vs anesthetized), approaching even closer to establishing the link between the mental and the physical worlds.

References

Aru J, Bachmann T, Singer W, Melloni L (2012) Distilling the neural correlates of consciousness. Neurosci Biobehav Rev 36:737–746. CrossRef

Balduzzi D, Tononi G (2008) Integrated information in discrete dynamical systems: motivation and theoretical framework. PLoS Comput Biol 4:e1000091. CrossRef

Balduzzi D, Tononi G (2009) Qualia: the geometry of integrated information. PLoS Comput Biol 5:e1000462. CrossRef Medline

Baron F, van Kempen J, Kawasaki H, Kochvah CK, Oya H, Howard MA, Adolphs R, Tsuchiya N (2017) Intracranial markers of conscious face perception in humans. NeuroImage 162:322–343. CrossRef

Barrett AB (2014) An integration of integrated information theory with fundamental physics. Front Psychol 5:63. CrossRef Medline

Barrett AB, Seth AK (2011) Practical measures of integrated information for time-series data. PLoS Comput Biol 7:e1001052. CrossRef Medline

Barrett AB, Barnett L, Seth AK (2010) Multivariate Granger causality and generalized variance. Phys Rev E Stat Nonlin Soft Matter Phys 81:41907. CrossRef

Bialek W, Nemenman I, Tishby N (2001) Predictability, complexity, and learning. Neural Comput 13:2409–2463. CrossRef Medline

Brainard DH (1997) The psychophysics toolbox. Spat Vis 10:433–436. CrossRef

Brenner N, Strong SP, Koberle R, Bialek W, Stevenick RR (2000) Synergy in a neural code. Neural Comput 12:1531–1552. Medline

Chang L, Tsao DY (2017) The code for facial identity in the primate brain. Cell 169:1013–1028.e14. CrossRef Medline

de Graaf TA, Hsieh-P-J, Sack AT (2012) The “correlates” in neural correlates of consciousness. Neurosci Biobehav Rev 36:191–197. CrossRef

Freiwald WA, Tsao DY, Livingstone MS (2009) A face feature space in the macaque temporal lobe. Nat Neurosci 12:1187–1196. CrossRef Medline

Gómez C, Hornero R, Abásolo D, Fernández A, Escudero J (2007) Analysis of the magnetoencephalogram background activity in Alzheimer’s disease patients with auto-mutual information. Conf Proc IEEE Eng Med Biol Soc 08:239–247. CrossRef

Grill-Spector K, Knouf N, Kanwisher N (2004) The fusiform area subserves face perception, not generic within-category identification. Nat Neurosci 7:555–562. CrossRef

Haxby JV, Hoffman EA, Gobbini MI (2000) The distributed human neural system for face perception. Trends Cogn Sci 4:223–233. CrossRef

Henrikssoon L, Mur M, Kriegeskorte N (2015) Factoty—A face-feature map with face-like topology in the human occipital face area. Cortex 72:156–167. CrossRef Medline

Hoel EP, Albantakis L, Marshall W, Tononi G (2016) Can the macro beat the micro? Integrated information across spatiotemporal scales. Neurosci Conscious 2016:niw012. CrossRef

Hohwy J (2013) The predictive mind. Oxford: Oxford University Press.

Joë J, Lamme VAF (2005) Repression of unconscious information by conscious processing: evidence from affective blindsight induced by transcranial magnetic stimulation. Proc Natl Acad Sci USA 102:10747–10751. CrossRef

Julitta B, Vallverdu M, Melia USP, Tupaika N, Josip M, Jensen EW, Struys MM, Vereecke HE, Caminal P (2011) Auto-mutual information function of the EEG as a measure of depth of anesthesia. Conf Proc IEEE Eng Med Biol Soc 2011:2574–2577. CrossRef

Koch C (2004) The quest for consciousness. Englewood: Roberts & Company Publishers.

Kriegeskorte N, Mur M, Bandettini P (2008) Representational similarity analysis connecting the branches of systems neuroscience. Front Syst Neurosci 2:4.

Lord L-D, Stevner AB, Deco G, Kringelbach ML (2017) Understanding principles of integration and segregation using whole-brain computational connectomics: implications for neuropsychiatric disorders. Phil Trans R Soc A 375:20160283. CrossRef

Marr D (1982) Vision: a computational investigation into the human representation and processing of visual information. New York: Henry Holt and Co. Inc.

Oizumi M, Ishii T, Ishibashi K, Hosoya T, Okada M (2010) Mis-matched decoding in the brain. J Neurosci 30:4815–4826. CrossRef

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

Oizumi M, Amari S, Yanagawa T, Fuji I, Tsuchiya N (2016a) Measuring integrated information from the decoding perspective. PLoS Comput Biol 12:e1004654. CrossRef

Oizumi M, Tsuchiya N, Amari S (2016b) Unified framework for information integration based on information geometry. Proc Natl Acad Sci USA 113:14817–14822. CrossRef

Parvizi J, Jacques C, Foster BL, Withoff N, Rangarajan V, Weiner KS, Grill-Spector K (2012) Electrical stimulation of human fusiform face-selective regions distorts face perception. J Neurosci 32:14915–14920. CrossRef

Pegna AJ, Khateb A, Lazeyras F, Seghier ML (2005) Discriminating emotional faces without primary visual cortices involves the right amygdala. Nat Neurosci 8:24–25. CrossRef

Pelli DG (1997) The VideoToolbox software for visual psychophysics: transforming numbers into movies. Spat Vis 10:437–442. CrossRef

Puce A, Allison T, McCarthy G (1999) Electrophysiological studies of human face perception. III: effects of top-down processing on face-specific potentials. Cereb Cortex 9:445–458. CrossRef

Ramsøy TZ, Overgaard M (2004) Introspection and subliminal perception. Phenomenol Cogn Sci 3:1–23. CrossRef

Rangarajan V, Hermes D, Foster BL, Weiner KS, Jacques C, Grill-Spector K, Parvizi J (2014) Electrical stimulation of the left and right human fusiform gyrus causes different effects in conscious face perception. J Neurosci 34:12828–12836. CrossRef

Reimann MW, Nolte M, Scolamiero M, Turner K, Perrin R, Chindemi G, Dlotko P, Levi R, Hess K, Markram H (2017) Cliques of neurons bound into cavities provide a missing link between structure and function. Front Comput Neurosci 11:48. CrossRef

Schäfer J, Strimmer K (2005) A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. Stat Appl Genet Mol Biol 4:Article32. CrossRef

Tong F, Nakayama K, Vaughan JT, Kanwisher N (1998) Binocular rivalry and visual awareness in human extrastriate cortex. Neuron 21:753–759. CrossRef

Tononi G (2004) An information integration theory of consciousness. BMC Neurosci 5:42. CrossRef Medline

Tononi G (2012) Integrated information theory of consciousness: an updated account. Arch Ital Biol 150:56–90. CrossRef

Tononi G, Boly M, Massimini M, Koch C (2016) Integrated information theory: from consciousness to its physical substrate. Nat Rev Neurosci 17:450–461. CrossRef

Tsuchiya N (2017) “What is it like to be a bat?”—a pathway to the updated account. Proc Natl Acad Sci USA 115:8062–8067. CrossRef

Tsuchiya N, Koch C (2005) Continuous flash suppression reduces negative afterimages. Nat Neurosci 8:1096–1101. CrossRef

Tsuchiya N, Taguchi S, Saigo H (2016) Using category theory to assess the relationship between consciousness and integrated information theory. Neurosci Res CrossRef

September/October 2017, 4(5) e0085-17.2017 eNeuro.org