**Rationale and Computation of Integrated Information (Φ)**

From an information-theoretic perspective information is the reduction of uncertainty (Shannon, 1948). In turn, information may be quantified by considering how a system in its current state $S\_{0}$ constrains the system’s potential past and future states. Figure S1 illustrates this principle form within the purview of $C$ at time $t$ for the system with an XOR gate. Under this scenario, if $C$ is currently active, then at time $t$-1 by necessity either A was active, B was active, A and C were active, or B and C were active (Figure S1, left panel). The probability distribution of past states that could have been causes of $C=1$ is its cause repertoire $p\left(C=1\right)$. On the other hand, if it is unknown in what state $C$ is in, $t$-1 is unconstrained $p^{uc}(ABC^{past})$. A similar rationale applies to future states wherein the current state of $C$ constrains its future potential states, and the effect repertoire is thus the probability of being in any given state given that $C$ is current active, or $p\left(C=1\right).$ The amount of information that $C=1$ specifies about the past is its cause information (CI) and the amount it specifies about the future is its effect information (EI). CI and EI are respectively measured as follows,

$$CI=EMD(p\left(C^{t}=1\right) || p^{uc}(ABC^{t-1}) Eq.1$$

$$EI=EMD(p\left(C^{t}=1\right) || p^{uc}(ABC^{t+1}) Eq.2$$

where $EMD$ refers to earth mover’s distance (Rubner et al., 2000), the minimal cost of reshaping one distribution (e.g., unconstrained) into the other (e.g., constrained) or area of distribution moved times the distance moved. Finally, the total amount of cause-effect information (CEI) specified by $C=1$ is the minimum value between CI and EI. This results from the fact that both CI and EI may act as limiting cases – an information bottleneck – and hence minimize the CEI of the system as a whole (see Oizumi et al., 2014, for detail). Finally, while CEI measures information, the IIT conjectures that consciousness is *integrated* information. That is, information generated by the system above and beyond that generated by its constituent parts. Hence, the system as a whole is iteratively partitioned into all possible subsystems or purviews and the process delineated above is evaluated for each of these components. Similar to CEI, integrated information is calculated as the $EMD$ between the cause-effect repertoire specified by the system as a whole and the cause-effect repertoire of the partitioned system. Φ is the distance between the system as a whole and the system-partitioned that makes the least difference; the minimum information partition. That is, Φ is the degree to which the cause effect repertoire for the system as a whole differs from the next most informative partition.



**Figure 1-1. Illustration of cause and effect repertoires and the constraints imposed on potential probability distributions by the fact that C=1.** Cause (left) and effect (right) repertoires for a system with three nodes as the one illustrated in Figure 1, and as a function of whether the past-future is constrained to C=1 (top) or not (bottom).

Information integration (phi, Φ) was calculated for the multisensory convergent and integrative networks using the transitions probability matrices illustrated in Figure 2-1 (see below) and as implemented in *PyPhi* (Mayner et al., 2017)with Python 3.4.

**Formalizing the Role of Multisensory Integration in Consciousness**

To formally ascertain the putative role multisensory integrative (vs. convergent) neurons within a network in bearing consciousness (according to the IIT), we built two biologically inspired simple neural networks (Figure S2A). These networks each have 3 nodes, two of which may be considered analogous to unisensory areas (nodes A and B) and a third (node C), which receives projections from the unisensory areas and may be considered analogous to a multisensory area. As is well established in biological systems, the multisensory area equally projected back to unisensory areas (Bizley et al., 2007, Cappe et al., 2009, Ghazanfar and Schroeder, 2006). The two networks were identical with exception that for one network (Figure S2A, left panel) node C was an “XOR” gate, while for the other it was an “AND” gate (Figure S2A, right panel). The XOR gate results in a logical “true” (or ‘1’/ ‘HIGH’) when the number of true inputs is odd. In this case, given the system architecture, the node C would be active if on the previous time step one and only one of gates A or B was active. Thus, node C can in principle be active following activity in either node A or B, but importantly does not respond preferentially when both are active. On the other hand, the other network, where node C is an “AND” gate, responds exclusively when on the precedent time-step both gates A and B were active. That is, this gate most faithfully mimics the behavior of integrative multisensory neurons that may or may not overtly respond indiscriminately to distinct sensory inputs, but importantly are most driven by spatio-temporally coincident multisensory inputs. Hence, the network formed with an XOR gate (Figure S2A, left) instantiates a network with neurons that are indiscriminant to the nature of sensory input, while in contrast the network formed with an AND gate (Figure S2A, right) instantiates a network with neurons that integrate sensory information, i.e., responds non-linearly to the addition of sensory stimuli from distinct modalities (Stein and Stanford, 2008, Wallace et al., 2006). The architecture of these systems dictates the composition of transition probability matrices (TPMs), which guides transitions between states (i.e., neurons that are ‘active’ at different time-points). In Figure S2B these TPMs have been depicted (left and right respectively for the multisensory convergent and multisensory integrative systems) and highlighted for their distinctive features. Namely, in the case of the convergent network, when ABC nodes are in state 100 (respectively, A, B, and C) or 010 (green rows), activation of the multisensory node will follow. This is not true if the convergent network is in state 110 (red row). The opposite is true for the network that integrates. Given these TPMs, Φ can be calculated when the state of the network is ABC = 001 (multisensory node active). Results indicate that in fact a network with a node with integrative capacity in principle may bear a greater degree of consciousness (Φ = 0.78) than one that simply responds indiscriminately to stimuli from distinct sensory modalities (Φ = 0.25).

****

**Figure 2-1. Formalizing the role of multisensory integrative neurons in bearing consciousness according to IIT. A)** Depiction of a multisensory convergent (left) and integrative (right) network. There is no connection between A and B nodes, as these transition probability values are zero. The dashed connections leading to A and B are to illustrate that these putative unisensory areas receive input from downstream neural areas, yet they play no role in the ITT-driven model. **B)** The transition probability matrices (TPM) for a deterministic (e.g., probability is either 0 or 1) convergent (left) and integrative (right) network are illustrated. State t is represented in the abscissa and t+1 on the ordinate. Green and red rows are highlighted to emphasize key difference between the convergent and integrative networks, yet these differences are not exhaustive (however do dictate the rest of differences). **C)** The Φ value associated with the convergent (left) and integrative (right) TMPs as determined in *PyPhi* (Mayner et al., 2017).

**S1 and vPM Firing Rates as a Function of Sensory Modality and Awareness**

Regarding the firing rate, analyses on non baseline-corrected activity indicated a clear generalized decrease in firing rate when monkeys were rendered unconscious (p<0.01 at all time points; Aware; M = 4.43 spikes/s, S.E.M = 0.008 spikes/s; Unaware; M = 2.44 spikes/s, S.E.M = 0.007), a significant difference in spiking activity across the areas recorded between 50ms and 160ms post-stimuli onset (p<0.01, S1, M = 5.68 spikes/s, S.E.M = 0.008 spikes/s; vPM, M = 4.88 spikes/s, S.E.M = 0.006 spikes/s), and a significant main effect of stimulation type (i.e., AT, T, A, N) between 50ms and 480ms post-stimuli onset (AT, M = 4.14 spikes/s, S.E.M = 0.01 spikes/s; T, M = 4.31 spikes/s, S.E.M = 0.01 spikes/s; A, M = 3.75 spikes/s, S.E.M = 0.006 spikes/s; N, M = 3.28 spikes/s, S.E.M = 0.001 spikes/s). Importantly, a stimulation-type (i.e., AT, T, A, N) by area recorded (i.e., S1 vs. vPM) interaction was also evident (p<0.01, between 60ms and 210ms post-stimuli onset), driven by the fact that vPM responded to A stimulation (M = 3.21 spikes/s, S.E.M = 0.10 spikes/s), while S1 did not (M = 2.18 spikes/s, S.E.M = 0.10 spikes/s). Thus, in sum and as expected, these analyses demonstrated that propofol silenced spiking activity generally, that neurons in S1 and vPM responded differently to distinct sensory stimuli between 50 and 480ms post-stimuli onset, and that vPM but not S1 responded to auditory stimulation. The baseline-corrected analyses (depicted in Figure 2, rows 1 and 3) largely demonstrated analogous results, while indicating that the bifurcation in evoked responses (as opposed to baseline responses, as indicated above) between states of consciousness occurred (regardless of sensory modality) 80 ms post-stimuli onset (p<0.01, averaged across 80-1000ms post-stimuli onset; Aware, M = 0.48 spikes/s, S.E.M = 0.04; Unaware, M = 0.09 spikes/s, S.E.M = 0.01 spikes/s) and also highlighting a consciousness state by sensory modality (p<0.01 between 40-110ms, and 200-380ms) as well as 3-way (modality, state, and area) interaction (p<0.01, 410-610ms post-stimuli onset). The time-periods demonstrating a significant difference in evoked activity as a function of state of consciousness are shaded in gray in Figure 1 (main text) separated by area recorded and sensory stimulation, while the time-periods demonstrating a significant response vis-à-vis baseline are indicated by horizontal lines in each panel (see Figure 1).

**S1 and vPM Fano Factors as a Function of Sensory Modality and Awareness**

Regarding fano factors (FFs), results demonstrated heightened variability under unaware (M = 1.45, S.E.M = 7.3e-4) than aware (M = 1.16, S.E.M = 5.6e-4) conditions (see Ecker et al., 2014 for a similar result), while both of these were significantly greater than 1 (unaware, p < 4.76e-92; aware, p = 4.76e-92), and hence likely exhibiting inter-trial variability above and beyond what is presumed to be attributable intrinsically to neurons (i.e., FF = 1). Similarly, FFs were larger in S1 (M = 1.32, S.E.M = 6.91e-4) than vPM (M = 1.22, S.E.M = 5.70e-4), throughout the post-stimuli period (p<0.01, for exemption of the period between 80ms and 120ms post-stimuli onset. The period between 50ms and 270ms post-stimuli onset demonstrated a significant difference in FFs as a function of stimulus modality (p<0.01), with the AT (M = 1.29, S.E.M = 0.02) and T (M = 1.28, S.E.M = 0.03) conditions being the less variable (AT vs. T, p = 0.58) than the A (M = 1.31, S.E.M = 0.02) and N (M = 1.33, S.E.M = 0.02) conditions (T vs. A, t = 2.03, p = 0.041; A vs. N, p = 0.43). Importantly, FFs also demonstrated a consciousness state by recording area interaction (p<0.01, between 200ms and 320ms post-stimuli onset) and a recording area by stimulus modality interaction (p<0.01, between 50ms and 180ms post-stimuli onset). The latter interaction was driven by a main effect of stimuli modality that was sustained in S1 (p<0.01, between 50ms and 250ms, as well as 350ms and 540ms post-stimuli onset) and only transient in vPM (p<0.01, between 110 and 140ms post-stimuli onset), while the former is attributable to a rapprochement in FF between consciousness states in S1 and not in vPM. Indeed, this last effect is further appreciable when correcting FFs for baseline (Figure 1) as a further quenching in variability in S1 (vs. vPM) specifically to AT and T sensory stimulation (contrasts between aware and unaware conditions; S1; AT, p<0.01 between 160ms-200ms, T, p<0.01 between 180ms-220ms, never for A and N; vPM, never). As for firing rates, the time-periods demonstrating a significant difference in FF as a function of state of consciousness are shaded in gray in Figure 1 (main text) separated by area of recording and sensory stimulation type. Time-periods demonstrating a significant quenching in FF vis-à-vis baseline are indicated by horizontal lines in each panel (see Figure 1).

**Lempel-Ziv Complexity as a Function of Sensory Modality and Awareness**

Figure 5A illustrates normalized LZ (see Andrillon et al., 2016, Noel et al., 2018, and Methods), both in its baseline-corrected and non-corrected format, and as a function of consciousness state (aware = colored; unaware = gray) and sensory stimulation. Regarding the non-corrected values, a 2 (consciousness state; aware vs. unaware) x 2 (recording area; S1 vs. vPM) x 4 (stimulation type; AT, T, A, N) ANOVA most strikingly revealed that unaware states (M = 0.88, S.E.M = 0.003) were generally more complex (p<0.01 at all time-points) than aware states (M = 0.81, S.E.M = 0.004). This analysis also revealed a main effect of recording area between 50ms and 100ms post-stimuli onset (p<0.01), as well as a main effect of stimulation between 50ms and 250ms (p<0.01). This analysis equally indicated a significant interaction between recording area and stimulation type (p<0.01 between 50ms and 150ms post-stimuli onset). The interaction was driven by a significant main effect of stimulation that lasted longer (p<0.01, between 50ms and 250ms post-stimuli onset) in S1 than vPM (p<0.01 between 100 and 150ms). Once normalized LZ was corrected for baseline, analyses specified a main effect of consciousness state specifically between 200 and 400ms post-stimuli onset (p<0.01), indicating that not only was overall LZ different across consciousness states, but the evoked nature of this measure equally differed. This main effect was driven by the AT and T conditions, where complexity returned to it’s baseline value more readily under unaware (AT, and T, return to baseline at 300ms) than aware states (AT and T, return to baseline at 350ms). The rest of statistical contrasts followed the same pattern as for the non-corrected values. The time-periods demonstrating a significant difference in evoked activity as a function of state of consciousness are shaded in gray in Figure 5A separated by area recorded and sensory stimulation, while the time-periods demonstrating a significant response vis-à-vis baseline are indicated by horizontal lines in each panel (see Figure 5A). In sum, therefore, the state of awareness is seemingly indexed in spiking activity by an overall lower level of LZ complexity (see Figure 5A, non-corrected normalized LZ), as well as by a more sustained negative deflection evoked by sensory stimulation (see Andrillon et al., 2016, for a similar observation).

**Percentage of Trials Evoking Significant Firing in S1 and vPM as a Function of Sensory Modality and Awareness**

The percentage of trials that resulted in significant firing of S1 neuron was altered by the sensory modality of the stimuli presented and the consciousness level of the monkey. In particular, results demonstrated a main effect of consciousness state (Z=1294, p<0.001), stimulation modality (χ2=51.52, p<0.001), and an interaction between these variables (χ2=80.99, p<0.001). The interaction was driven by a significant main effect of stimuli type during consciousness (χ2 =91.18, p<0.001), but not unconsciousness (χ2=4.07, p=0.19). Regarding significant activation of pre-frontal cortex neurons, once again results demonstrated further activation consciously (M=6.9%) than unconsciously (M=3.1%; Z=1319, p<0.001), a main effect of stimulation type (χ2=105.7, p<0.001), and an interaction between these variables (χ2 = 233.11, p < 0.001). The interaction was driven by differential trial-activation percentages as a function of stimulation type when the monkeys were conscious (χ2 =133.7, p<0.001) but not unconscious (χ2= 7.51, p=0.08).