Consciousness as Integrated Information: A Provisional Philosophical Critique

Anthony F. Peressini
Marquette University, anthony.peressini@marquette.edu

Consciousness as Integrated Information: a Provisional Philosophical Critique*

DRAFT: PLEASE DO NOT CITE

Anthony Peressini
Department of Philosophy
Marquette University
September 2012

email: anthony.peressini@marquette.edu

Abstract:

Giulio Tononi (2008) has offered his integrated information theory of consciousness (IITC) as a “provisional manifesto.” I critically examine how the approach fares. I point out some (relatively) internal concerns with the theory and then more broadly philosophical ones; finally I assess the prospects for IITC as a fundamental theory of consciousness. I argue that the IITC’s scientific promise does carry over to a significant extent to broader philosophical theorizing about qualia and consciousness, though not as directly as Tononi suggests, since the account is much more focused on the qualitative character of experience rather than on consciousness itself. I propose understanding it as “integrated information theory of qualia” (IITQ), rather than of consciousness.

1. Consciousness as Integrated Information

Giulio Tononi (2008) has recently offered his integrated information theory of consciousness (IITC) as a “provisional manifesto.” I critically examine how the approach fares. I point out some (relatively) internal concerns in section 2 and then in section 3 some more broadly philosophical ones, and finally in section 4, I assess the prospects for integrated information (II) as a fundamental theory of consciousness. I argue that the
IITC's scientific promise does carry over to a significant extent to broader philosophical theorizing about qualia and consciousness, though not as directly as Tononi suggests, since the account seems much more focused on the qualitative character of experience rather than on consciousness itself.

The formal definition for the amount of integrated information in a system, \( \Phi \), depends on the notion of relative entropy from modern information theory. Given a system, \( X \), characterized by mechanism \( mech \), where \( mech \) consists of \( n \) discrete states, \( x_1, x_2, \ldots x_n \), one considers the probability distribution of its possible states, \( p(X(mech))=\{p_1, p_2, \ldots p_n\} \).

One important distribution for \( X(mech) \) is its maximum entropy (equivalent to uniform probability in many simple systems) distribution: \( p(X_0(\text{maxH})) \), where the subscript 0 indicates time \( t=0 \), and would, for example, look like \( \{\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4} \} \), for \( n=4 \). The maximum entropy distribution is the “zero point” from which one measures information as “distance.”

The distance measure between two probability distributions, \( p \) and \( q \), is given by the Kullback-Leibluer divergence \( H \), (also know as relative entropy), defined as: \( H[p|q] = p_i \log(p_i/q) \). Tononi then defines the effective information \( (ei) \) generated by a mechanism in a particular state \( x_1 \) at \( t = 1 \) as

\[
\text{ei}(X(mech, x_1)) = H \left[ p(X_0(mech, x_1)) \parallel p(X_0(\text{maxH})) \right],
\]

where this is to be understood as the information generated by the system’s mechanism and state at \( t=1 \) about the system’s prior state at \( t=0 \). Finally the integrated information \( \Phi \), for a system \( X \) in state \( x_1 \) is the difference (measured by relative entropy) between the probability distribution generated by the system as a whole, \( p(X_0(mech, x_1)) \) and the
probability distribution generated by $X$'s decomposition into parts that leave the minimal
information unaccounted for, denoted MIP:

$$\Phi(X(mech, x_1)) = H \left[ p(X_0(mech, x_1)) \right| \Pi p(kM_0(mech, \mu_0)) \right], kM_0 \text{ ranges over MIP.}$$

Heuristically, this difference can be thought of as the information of the system not
accounted for by its parts.

Having set out this formal machinery, Tononi expresses his provisional manifesto as
follows: the IITC proposes that consciousness is II, specifically that (1) the quantity of
consciousness corresponds to the amount of II generated by a complex of elements, $\Phi$, and
(2) the quality of experience is specified by the set of informational relationships generated
within that complex. The informational relationships are formally characterized by the
properties of solids (polytropes) in the appropriate $2^n$ dimensional space, where $n$ is the
number of states in the mechanism. This space is called Q-space and the $2^n$ dimensional
solids are generated by how the probability distributions for the system change as a
function of the connections/transitions in the mechanism. Tononi (2008, esp 224ff)
develops many formal aspects of Q-space and shows how these neatly mirror pre-theoretic
conceptions of the qualitative aspect of experience. For example, he develops a property of
the q-arrows (sides) of solids in Q-space he calls entanglement. This formal property is
argued to capture the notion of modes in qualitative experience (sight, sound, etc.).

Another important feature is that qualia so construed in Q-space are context dependent in
a way that parallels our sense that the subjective qualities of experience (qualia) are
dependent on the broader “qualia context,” e.g., the particular red qualia a subject
experiences when viewing the red of a stop sign may well be different if it is part of a visual
field that includes a high percentage of reds of varying shades. Finally, some shapes in Q-
space seem to be elementary in that they cannot be further decomposed (do not contain any more densely entangled sub-sub-modes), and these would seem to correspond to “what philosophers call a ‘quale’ in the narrow sense—say a pure color like red, or a pain, or an itch…” (230).

To sum up the account, (in Tononi’s words): “Perhaps the most important notion emerging from this approach is that an experience is a shape in $Q$. According to the IITC, this shape completely and univocally specifies the quality of the experience” (228) [original italics]. I move on now to an appraisal of the IITC approach.

2. Internal Concerns

In this section I present concerns with the II approach that have to do with the internal cogency of II as a theory of a fundamental property of the brain. These concerns are whether the notion of II is well defined in a formal sense for an arbitrary system and second how the (notorious) difficulty in interpreting probability affects the II account.

2.1 Is Integrated Information Well Defined?

The notion of a good definition in mathematics (a term being well defined) means roughly that the defined operation or concept or entity is unambiguous. For example, with cosets of a group relative to a subgroup $H$, one defines the right coset as: $Ha=\{h \epsilon H \mid h \cdot a\}$. A product operation on right cosets is then defined as $Ha \cdot Hb = H(a \cdot b)$, but one has to establish that the result is unambiguous. In this case that amounts to showing that the coset picked out by $Ha$ doesn’t depend on the particular element, $a$, used to represent it. Thus one must prove that if $b \epsilon Ha$, then $Ha = Hb$. The concern here is that $\Phi$, which is to
be understood as the amount of consciousness of a complex, be well defined in that the value of $\Phi$ for a given system must always exist and be unique. Now of course for the simple systems of 4 nodes that are considered by Tononi, $\Phi$ is a straightforward calculation, but things aren’t always that simple.

Recall that the definition of $\Phi$ for a system $X$ in a state $x_1$ as characterized by mechanism $\text{mech}$ is the difference between the probability distribution generated by the system as a whole, $p(X_0(\text{mech}, x_1))$ and the probability distribution generated by $X$’s decomposition into parts that “leaves the least information unaccounted for,” (220), denoted MIP (minimum information partition). The relative entropy of the system measures this difference between the II of the whole and its maximally “informative” partition, again formally:

$$\Phi(X(\text{mech}, x_1)) = H \left[ p(X_0(\text{mech}, x_1)) \right] - \Pi \left[ H \left( p(M_0(\text{mech}, \mu_0)) \right) \right],$$

where $M_0$ ranges over MIP. Thus the definition of $\Phi$ depends on identifying a particular decomposition (MIP) from the set of all possible decompositions, but of course there are in general many different ways of dividing a system into parts. Tononi adds a further constraint on the set of “admissible” decompositions: we understand the eligible decompositions as being determined by the causal mechanism of the system (220). The concern is that this still need not determine a unique decomposition. In order for $\Phi$ to be well defined, it must be possible to show that the set of (causally) possible decompositions, $D$, has a minimal element relative to the value in question, in this case the “distance” from $p(X_0(\text{mech}, x_1))$. Regardless of what is to be minimized (or maximized), the question of whether such a min/max exists (and is unique) over $D$ is unclear. Unless, such a result can be established, the application of the formalism comes apart. And the deeper consequence
of this (assuming no proof is forthcoming) is that a complex has no definite $I$, and if $I$ is consciousness, then the complex has no definite amount of consciousness—nor is there a definite answer to whether it has consciousness at all (whether $\Phi > 0$).

Even if one had reason to believe that a unique $\Phi$ exists in general, further questions remain. Why think that the degree of consciousness of the complex should track the $I$ associated with the minimal information partition (MIP)? It is true that on Tononi’s account, $I$ is defined in terms of the MIP, so strictly speaking there no $I$ except as calculated in terms of the MIP. Nonetheless one may consider notions closely related to $I$ (defined relative to other significant partitions) that may correspond equally well to the complex’s degree of “informational integration,” understood more generally. In other words, consciousness may align with “informational integration” and yet not track the state’s entropy relative to the minimal information partition ($I$ defined as $\Phi$), but rather relative to some other partition favored for some physically or informationally significant reason.

Finally, there is the assumption (built in from the beginning) that the system is characterized by “a certain mechanism.” This would seem to limit the analysis to systems that are (or can be approximately or provisionally characterized) mechanistically, which brains systems may or may not be. Even if $I$ (understood more generally as above) is the key to consciousness, must it be characterized mechanistically? By relativizing $I$ to the $\Phi$ of a system under a particular (mechanistic) description, the approach is again susceptible to “well definition” problems: on what principle does one adjudicate between potentially distinct mechanistic characterizations of system, each with distinct $\Phi$’s? The amount consciousness of a system is presumably a definite value and yet the system has the
potential to have multiple $\Phi$ values, depending on what particular (mechanistic) description one uses. Tononi does have a response to this concern in that he seems to maintain that the brain (and all physical systems) have a fundamental mechanistic description that is uniquely determined by its physical composition, namely, the one with “minimal elements and minimal interactions—as elementary as they come in terms of physical implementation” (2008, 234). If sustainable, such a view would alleviate some of these last concerns. There is, however, reason to be skeptical of this reductive mechanistic assumption, not the least of which is that it seems to be in tension with Tononi’s recognition that II is generated at “multiple spatiotemporal scales” and that one should vary the grain size in order to maximize $\Phi$ (2008, 235-36).

It is a limiting feature of the II approach (as developed via $\Phi$) that it requires the system under consideration to be mechanistic. Simple node networks with on/off connections are not without their predictive and investigative use, but if the II approach is really limited to mechanistic understandings of the brain, then it looses a great deal of its plausibility as the key to consciousness. It seems it might be possible to generalize the II approach in some directions. For example, the restriction of the formal characterizations of the systems to (finite) discrete states is not a mathematical necessity; the formal mathematics of relative entropy for continuous random variables was worked out by Kolmogorov (1956) and has been since refined with ancillary results (see Cover and Thomas 2006). The problem with this generalization of II is that while the quantity of consciousness is preserved, it isn’t at all clear that the quality of consciousness (as developed by way of Tononi’s Q-space) can be. The dimension of the Q-space for a given $X(mech,x_1)$ is the number of possible states for the mechanism, which in the case of a
continuous variable is infinite. While infinite dimensional spaces are well understood in some cases (i.e., those with finite norms or distance functions known as Banach spaces), it isn’t clear how Q-spaces will fare in infinite dimensions. The points in Q-space would be functions, and the maximum entropy distribution would again be the baseline for understanding the various properties (divergence, null context, distance between distributions, etc.), but it isn’t a given in this context that the maximum entropy distribution would exist (be finite). So again one is placed in a position to establish (as in the case of the MIP) significant mathematical results in order for the formalism to be mathematically rigorous and have meaningful physical interpretations. Tononi (2008, 229) acknowledges something like this.

2.2 Probability and Integrated Information

The difficulties with interpreting probability go to back to the very beginning of the formal probability calculus. I am sympathetic with the idea that probability is often implicitly defined in its scientific contexts and that a detailed interpretation is not something a scientist working on a scientific problem need have in hand. Also, a pluralist account of probability is likely correct, so pressing a particular interpretation against the II account would be unreasonable, or at least premature. Nonetheless, since this scientific theory of consciousness is being offered as a solution to the philosophical problem of consciousness (and also because consciousness, unlike say population genetics, has from the start a “foot planted deeply in the mud” of philosophy), the worry about the interpretation of the probability is germane.
The probability involved in information theory comes from the definition of entropy and thereby effective information, \((ei)\). Recall that \((ei)\) involves probability distributions of the form \(p(X_0(mech, x_1))\), that is, a probability distribution over all possible states, \(x\), of the system \(X_0\), specifying the probability that the system’s prior state was \(x\), where the system \(X_0\) is being considered as a mechanism specified by \(mech\) in state \(x_1\). An immediate question is whether probability is simply a mathematical device for measuring consciousness or whether probability is in some sense constitutive of consciousness. Obviously the former case raises fewer questions, but even if II were not essentially probabilistic, questions of interpreting the formal probability involved in its operationalization may make a difference in how (or whether) we are inclined to understand the consciousness as II proposal. For example, if one adopts a subjectivist interpretation (e.g., Bayesian), then as Michael Beaton argues Tononi’s approach via \(\Phi\) (and it’s internal measure of information) is misguided as a means of fleshing out our pre-theoretic notions of integrated information.\(^{10}\)

If, however, consciousness is understood as constituted (at least in part) by the probability involved in its definition via \(\Phi\), then much more turns on how we understand this probability. Consider the example of fitness in evolutionary biology—how probability functions in its standard definition as “the probability of surviving and reproducing.” Here the probabilistic property is generally not understood as corresponding to a fundamental feature of the universe, but rather something reducible in an ontological sense (though perhaps supervenient on the properties and hence irreducible in other senses) to the “causal push and pull of the universe.” If the II theory’s move is to identify consciousness with probability understood something like this (again, though one that is oddly predictive
of past "states" as opposed to the future "states" of fitness), it is difficult to see it as addressing the (apparently) fundamental, novel, emergent character of consciousness in the sense intended by Tononi.

In Tononi’s presentation, one finds a few hints as to how he is thinking of probability. The mechanistic framework with finite states naturally suggests a classical/frequentist interpretation. In this case probabilities are considered the frequencies of favorable events relative to either all possible (classical) or actual (frequentist) outcomes. There is solid precedent for the classical interpretation in the scientific orthodoxy of the previous century, but “it is now widely agreed that there are insuperable difficulties” with it (Eells 1991, note 7, p. 35), so it cannot but give one pause regarding understanding consciousness in terms of it.

Alternatively, since consciousness is to be understood as a fundamental and really “out there in the universe” (objective) phenomenon, then a more natural interpretation might be the propensity account. On the propensity interpretation, probabilities are understood as physical dispositions (propensities) to behave in a certain way; these are grounded in the physical make up of the system and thus have the virtue of rendering probability objective. On the other hand, subjectivist accounts (Bayesian) have the advantage of being tied neatly to notions of rationality, which has a natural connection to consciousness. And Tononi does describe the probabilities involved as “constituting information . . . in the classical sense of reduction of uncertainty or ignorance” (220), which sounds like a subjectivist interpretation.11

Since all of these interpretations are grounded (to varying degrees and in varying ways) in sciences like physics and biology, one might argue that if probability is good
enough for these sciences, then why not for neuroscience? But again, the issue here is not whether philosophical problems with understanding probability pose a threat to the scientific theories (they don’t), but rather whether such problems pose a threat to an interpretation of the scientific theory as addressing the philosophical problem of consciousness (I suggest they do). In particular, the move to understand consciousness in terms of II is in important senses equivalent to understanding it in terms probability, so questions about probability become questions that remain about consciousness. Since among other things the objective reality of probability is not quite settled, the philosophical import of the proposal cannot be assessed, even along such coarse lines as whether it is real (objective) property.

Given the problems with all interpretations of probability, the hope of simply solving the philosophical problem of consciousness by analyzing it in terms information seems notably problematic. Despite this, I’ll argue below that its scientific merit does offer tools for making philosophical progress too.

3. Philosophical Concerns

I now examine the philosophical plausibility of the move to identify consciousness with II. I begin with questions concerning the distinction between consciousness and qualia. Next I consider a more modest “qualia as II” proposal (IITQ), and finally I reassess the prospects for regaining from IITQ the more sweeping “consciousness as II” (IITC) identity.
3.1 Consciousness vs. Qualia

The term “consciousness,” despite David Chalmers’ and others efforts, still engenders confusion above and beyond the natural philosophical disagreement over one’s own favorite account. Some such confusion is present in Tononi’s account, so pinning it down will help bring into relief both strengths and shortcomings of the II approach. I propose to understand consciousness along the line proposed by Chalmers (following Nagel), to wit, let the predicate $\text{SIL}(x) \equiv \text{“there is something it's like to be x.”}$ So according to our intuitive understanding, $\text{SIL}(\text{Chalmers})$ is true and $\text{SIL}$(my thermostat) is false. Notice that this predicate formally leaves open whether $\text{SIL}(x)$ entails that $x$ has any particular qualia. It also leaves open whether there is in fact any natural property in the world to which the predicate corresponds. The SIL focus puts the emphasis on the “experiencing” aspect of conscious experience and doesn’t rush to conflate the subjective aspects and the qualitative aspects of consciousness; I return to this shortly.

The notion of qualia goes back to Lewis (1929) originally and vaulted into prominence because it has been offered as a sticking point for functionalist and other physicalist accounts of mind. Qualia are understood to be the “qualitative aspects” of conscious experience, e.g., if one is seeing a stop sign then the way the red appears to one in his/her phenomenological field is a red quale. There are presumably qualia associated with all sense modalities and other states like emotions, moods, etc. Without getting into the many thorny areas of qualia or being insistent on any particular understanding, I note the following further features of qualia: they appear practically ineffable, non-relational, non-public, and immediately available to the subject. For notation, let $Q_i(x)$ denote “$x$’s
experience has an *i* qualitative aspect,” so $Q_{\text{red}}(\text{Chalmers})$ is true when Chalmers is looking at a stop sign under normal conditions.

Consciousness and qualia are often run together in some way, e.g., Chalmers in his “hard problem” paper (1995) moves between Nagel’s “smoothing it is like” (SIL) notion as applied to creatures and the phenomenological properties of mental states themselves (qualia). But it isn’t obvious that the problem of what makes an organism a subjective experiencer in Nagel’s sense (SIL-consciousness) and the problem of what makes a state a qualitative or phenomenological are so simply related. Others have realized this, for example, Leopold Stubenberg (1998) believes that it requires a 350+ page book to *argue* that to be conscious is to have qualia. More recently, Michael Beaton (2009), in working on consciousness and qualia (by way of introspection), concerns himself with the “phenomenal aspect of consciousness: with qualia; with the ‘something it is like’ to have an experience” (Beaton 2009, 88). Beaton is more careful, in a footnote (#1, p. 89) he writes that “qualia are the characteristic properties of phenomenological consciousness: something is a state of phenomenological consciousness if and only if it has such properties” but intends not to “completely ignore the many other aspects of present within the broader concept of ‘consciousness.’” The point is that the problem of subjective experience (SIL-consciousness) ought not be thought of as identical to the problem of qualia.

To sort this out, it helps to attend carefully to the distinction between creature consciousness and state consciousness, and also of course between “something it is like” consciousness and other less philosophically problematic kinds. I use the term SIL-consciousness to refer to the property of being a conscious *experiencer* in Nagel’s
“something it is like” sense. I employ it in a strong sense in that SIL-consciousness is not reductively assumed to be simply the sum total of a creature’s qualitative or phenomenal or p-conscious states at a given moment.16 That is, the question of the nature of a creature’s SIL-consciousness is not assumed to be the same as the question of the nature of the creature’s p-conscious states. To avoid confusion, I will avoid the term “p-conscious,” since it is variously used to refer to both (a) qualia or phenomenality, understood as properties of mental states and also to (b) being a subjective experienter, understood as a property of a creature as a whole not its individual mental states. I move back and forth between the terms “qualia” and “phenomenality” depending on context with the understanding that qualia are the phenomenological properties of mental states.

In Tononi’s II account, these relationships are not explicitly addressed, but some reasoned guesses are in order. Initially he describes consciousness in a way that suggests SIL-consciousness:

“[consciousness] is what vanishes every night when we fall into dreamless sleep and reappears when we wake up or dream. It is also all we are and all we have: lose consciousness and, as far as you are concerned, your own self and the entire world dissolves into nothingness” (216).

Yet more often, especially in the scientific core of the paper, “to be conscious” amounts to nothing more than “to have a vivid experience of red” (234), at which point it isn’t SIL-consciousness under consideration, but rather a qualitative aspect of a conscious experience, Qi(x).

A closer look at the structure of Tononi’s argument is helpful. The main argument for his claim that that “consciousness = II” is (1) that the quantity of II in the brain seems to
be correlated with the presence of consciousness (awake/asleep sense), and (2) the qualitative aspects of consciousness can be modeled by the formal model generated by projecting entropy distributions for each of the $n$ possible states onto $n$-space in such a way that the formal properties of this Q-space capture many qualitative properties of conscious experience. The empirical evidence marshaled for (1) consists of illustrating how certain parts of the brain that are high in II, when lesioned or when “off line” because of deep sleep or anesthesia, lead to the subject no longer being conscious. The empirical evidence used in (2) is more robust in that it involves showing how the formal properties of Q-space line up with intuitive and empirical features of qualia and what are presumed to be their neuro-anatomical bases, thus allowing for more diverse directions of confirmation and the unification of disparate phenomena, both from cognitive psychology and neurology.¹⁷

From this schematic of the argument, it is clear that Tononi is addressing $Q_i(x)$, and not SIL-consciousness in (2). The more difficult question is whether SIL-consciousness is being taken up in (1), that is, whether there is reason to believe that the quantity of “consciousness” being characterized in (1) via $\Phi$ might be, as Tononi seems to suggest, the quantity of SIL-consciousness. I believe here Tononi is making a mistake that many philosophers make in this context, namely, conflating SIL-consciousness of a creature and the qualitative properties of mental states. He might thus be (mis?)interpreted as implicitly proposing a reduction of SIL($x$) to all of the $Q_i(x)$ at the given time. In this case we understand II to track consciousness (SIL-sense) in virtue of the fact that the II-gradient in Q-space tracks the qualitative features $Q_i(x)$ of experience.

This cannot but disappoint those who understand the problem of consciousness along traditional philosophical lines—focusing on the SIL understanding—because it
unceremoniously eliminates or ignores the “hard problem” and the explanatory gap issues.\(^{18}\) The identification of SIL-consciousness with informational properties leaves unaddressed why and how it is that II rich neuronal complexes have something it is like to be them as opposed to nothing at all. Notice, however, that the II account does address the question of why it is that the neural basis of my bluish experience has the qualitative character it does as opposed to a different qualitative character or none at all.

Thus there is a better case to be made for Tononi’s contribution. The reasoning just described fails to satisfy philosophically not because it sidesteps the “hard problem” or explanatory gap but because it conflates SIL-consciousness and the phenomenality of mental states and in so doing obscures the difference between the subjective and the qualitative character of conscious experience. Understood in its best light, the II account shouldn’t be seen as directly addressing SIL-consciousness with its all-important subjective character, but rather the qualitative character of consciousness (qualia), understood as distinct from the subjective character.

To make this clear, one must refine the difference between SIL-consciousness and qualia (phenomenality), making use of the distinction between the qualitative and subjective characters of conscious experience. Loosely following Levine (2001) and more recently Kriegel (2009), notice that conscious experience has at least the following two aspects: a qualitative character and a “for-me” or subjective character. When I experience a clear blue sky, the experience has a q-character (qualitative) of “bluishness” and an s-character (subjective) of “being mine” in that intimate first person subjective sense.\(^{19}\) Note that in distinguishing the subjective and qualitative aspects of experience, I am not suggesting that subjectivity is something we “experience” independently of qualia. My
point is that subjectivity is not (necessarily) constituted by the qualitative aspect of experience, though our sense of our own subjectivity may well be qualitative. With this distinction in hand, we can understand SIL-consciousness, the question of what makes an entity an experiencer in Nagel’s sense, as fundamentally about the s-character of conscious experience, and questions about the nature of qualitative states (or “phenomenological”) as fundamentally about the q-character of conscious experience.

I do not intend to assert that SIL-consciousness involves only the s-character and that qualia involve only the q-character; further philosophical consideration might be helpful on this question but in the end empirical work will be required to decide the issue. I am arguing only that SIL-consciousness is not obviously the same as qualia (as I have outlined them) and that while SIL-consciousness may well be implicated in both the s-character and q-character, the question of qualia is essentially about q-character and in principle can be empirically investigated independently of s-character. It may turn out that SIL-consciousness (in us) is always associated with qualia, but again that is an empirical question. Conversely, one may question whether having qualia entails being SIL-consciousness, i.e., does Q_i(x) entail SIL(x)? This entailment is more difficult to leave empirically open because given the way I’ve laid things out, another way of putting it is: does having a qualitative experience entail being an experiencer? It would seem to—and this may pose a difficulty for understanding the II approach as an account of qualia and not necessarily of consciousness.

As I’ve argued above Tononi’s empirical argument for “II = consciousness” directly supports at most “II = qualia,” and not that “II = being an experiencer.” But if having qualia entails being an experiencer, then a proponent of the (refined) II account, which asserts
that II rich complexes have qualia, had better convince us either (a) that such II rich complexes are themselves (or are necessarily embedded in) experiencers, or (b) that we ought to give up the entailment from qualia to experience. At first glance neither (a) nor (b) look very promising. But I’ll argue below there is a way to understand (b) that renders “II = qualia” as a coherent empirical possibility. But first, does Tononi have a convincing case for (a)? I don’t think so.

The argument Tononi gives (217-219) to support the identification of consciousness (SIL-sense) with II consists of two thought experiments. The first is intended to suggest that one important difference between our seeing a light and a simple photodiode “seeing” a light is our ability to discriminate between many more alternatives, so “the [II] theory says that the more specifically one’s mechanisms discriminate between what pure light is and what it is not ... the more one is conscious of it” (218). But sheer ability to discriminate is not enough to allow for conscious experience as his second thought experiment illustrates: a camera consisting of a million independent binary photodiodes, which can from the outside be considered to have $2^{1,000,000}$ states and thus an astronomical ability to discriminate. Tononi goes on to argue that, “since its 1 million photodiodes have no way to interact, ... there is no intrinsic point of view associated with the camera chip as a whole” (219). The reasoning here is that the irreducible and holistic nature of the unity of our phenomenological experience supports the idea that conscious experience is (or tracks) II because in a similar way, integrated neural mechanisms high in II cannot be “disconnected” or reduced into components and retain their II.

While such “thought experiments” are effective at illustrating how II works and how it tracks the “richness of experience” in terms of information and the ability to discriminate
one state from many others, it is doubtful that anyone hesitant to accept “SIL-consciousness as II” for philosophical or pre-theoretical or intuitive or phenomenological reasons would find such thought experiments compelling. The first experiment, in associating greater ability to discriminate a thing from other things with being “more conscious ... of it,” at best taps intuitions relevant to “consciousness” in the sense of the quantity of awareness or attention (a-consciousness), not SIL-consciousness. And while the second integrated example does get at something more than a-consciousness with its focus on the holistic and irreducible nature of conscious experience, the parallel between this and II’s “holistic and irreducible” character is thin stuff on which to ground the claim that SIL-consciousness is integrated information, as opposed to the claim that SIL-consciousness is dependent on integrated information.21

Then how might one understand the “qualia as II” project in light of the pre-theoretic entailment from qualia to experience? With the s- and q-character distinction in hand, consider again the conscious experience of a full field blue wall—the qualitative visual experience of it consists of a q-character, it’s bluishness, but also an s-character, its “for me-ness.” It is what one might call the “subjective consequence condition” (SCC) that if a state has a subjective character, then it is the state of an experiencer. Thus, Q_{blue}(x) can be understood to entail SIL(x) via its s-character in particular, and not necessarily its q-character. Symbolically it might be put something like the following.

Definition of qualia: \[ Q_{blue}(x) \iff q_{blue}(x) \land s_{blue}(x), \]

Subj. Consequence Condition: \[ s_{blue}(x) \rightarrow SIL(x), \]

therefore \[ Q_{blue}(x) \rightarrow SIL(x). \]
Tononi’s revised II account can then be understood as an account of the q-character of experience. Is this an account of qualia? Well, yes and no. If qualia and subjective experience (SIL) are conflated then the II account wouldn’t seem to be an account of qualia. If however, the problem of qualia is understood as distinct from the problem of SIL-consciousness, with the former being fundamentally about experience’s q-character and the latter fundamentally about experience’s s-character, then the II account makes sense as scientific theory of qualia in the sense of experience’s q-character. Thus on the integrated information account of *qualia*, the explanatory gap is bifurcated and refined, but it is still there (as it should be). It bifurcates into the qualia or Q-question and the subjective or S-question. The Q-question is why does the neural basis of my bluish experience have the qualitative character it does, as opposed to a different qualitative character or none at all? The S-question is why does the neural information processing system that is my brain have something it is like to be me, as opposed to nothing at all. The integrated information account of *qualia* addresses the Q-question and not (directly) the S-question.

So why opt for the qualia understanding of the II account? I recap my two reasons offered above and give one more. First, the problem of what constitutes an experiencer seems much more at the heart of the consciousness question, both from a philosophical perspective and from a pre-theoretic intuitive perspective. Second, allowing SIL-consciousness to be conceptually distinct from qualia leaves the theoretical space more open for empirical investigation (decides less from the “arm chair”) and “allows” important scientific contributions (like Tononi’s) to be relevant to the philosophical problem. And finally, this bracketing move can be found in other cognitive research programs, for example memory, which has been productively studied in its own right despite the fact that
most manifestations of memory involve it being consciously or subjectively presented. Memory research brackets the s-character of conscious memory experience so that work on memory itself and on what makes it conscious (its s-character) can proceed relatively independently. In a similar way, the II account of qualia may be understood as having identified qualia with a feature of the brain, and despite its connection to SIL-consciousness, has begun to investigate and understand it in its own right. 23

3.2 Correlation or Identity

Once one carefully distinguishes between SIL-consciousness (hereafter just “consciousness”) and qualia in the setting of II, the theory’s plausibility (and scope) becomes clearer. As I outlined above, the II approach has its most compelling and direct support as an account of qualia as opposed to consciousness. That is, Tononi’s considerations suggest the identification of the qualitative aspects of conscious experience, $Q_i(x)$, with II, since the majority of the detailed scientific theorizing concerns qualia and its connection with II via Q-space. The connections between the formal properties of Q-space and what we know of the neurophysiological and phenomenological properties of qualia are compelling. But again, this confirmation is limited to qualia. I will return to the question of how consciousness might be related to II below, but for now I want to consider limiting Tononi’s II proposal to qualia (IITQ). A remaining question is whether the IITQ account is enough to convince us that II actually is qualia, as opposed to being merely a neural correlate of qualia. To see why the IITQ account has the potential to go beyond mere correlation consider first why resisting identity claims is so compelling in the case of the mind.
The resistance to recognizing brain properties as *identical* to mental properties like qualia or consciousness (or even more behaviorally analyzed states like belief or desire) is dependent upon the coupling of two things: (1) the pre-theoretic (apparent) independence of the conscious mental realm from the brain, which is based in the fact that “observation” fails to reveal a connection between the mind and brain, and (2) the high-level, coarsely-grained, and simple nature of the brute identifications typically involved in an identity claim between the conscious mind and brain (e.g., pain = c-fibers firing). What is more, condition (1) is reinforced by condition (2): when the theory embedding the identity claim requires one to countenance relatively simple, coarsely-grained, and high-level brute identities between the mental and the physical, it strengthens the sense that the mental properties are distinct and merely correlated with the physical.

The II account begins to undo this bind by offering a complex, fine-grained, multi-level theory of associations and compositional identities in support of the identification of qualia with II. Where it comes down to “brute identifications,” they are embedded in a rich theoretical and explanatory context, which serves as further warrant for the identification. Thus as an account of qualia only in the q-character sense, IITQ’s claim that the qualitative properties of experience are identical to the II properties of the brain meshes more smoothly with pre-theoretic sensibilities than less sophisticated or more thoroughgoing identity claims do. And again, the II account doesn’t reduce qualia to physical properties, but rather identifies them with the informational properties of the physical system, as Tononi (2008, 232) makes explicit in his assertion that they are not reducible to the neural mechanisms that “generate” them.
A distinct question one might have regarding IITQ concerns how to explain the predictive and explanatory power of II connectivity in characterizing normal and altered states of consciousness. If II is qualia and not consciousness, then why does II track consciousness? That is, the amount of II appears to be correlated with the “amount of consciousness.” A number of explanations are available: (1) qualia is highly correlated with consciousness; (2) a certain highly connected and/or highly centralized set of connections naturally high in II might be the key to consciousness; (3) II/qualia (loss of) may well be the key to unconsciousness (loss of consciousness) without being the key to consciousness (vs. non-consciousness).

The last possibility, the one that grows out of confusion between unconscious and non-conscious, is one of which Tononi and other researchers approaching consciousness from his angle have to be wary. While the loss of consciousness (various forms of altered unconsciousness) may give us clues about consciousness, simply contrasting what happens when one moves between consciousness and “a dreamless sleep” runs the risk of missing ground-level conditions for the possibility of consciousness itself (conscious vs. non-conscious). Assuming then that the IITQ is plausible and enjoys at least preliminary confirmation, I now return to the question of what it might tell us about consciousness itself.

3.3 From Qualia back to Consciousness

Granting that the question of (SIL-)consciousness and qualia, understood with the q- and s-character distinction in mind, are distinct and that Tononi’s considerations support (at most) q-character identification with II, we may still find ways to connect II and
consciousness. What might the link between "qualia as II" and "consciousness as II" be? In Tononi’s (2008, 217-219) phenomenological thought experiments he offers the following possibility (see the discussion above in 3.1 for details):

1. consciousness is nothing more than the ability ("intrinsic potential") to discriminate between certain states of affairs and "a multitude of others" [by way of the q-character of its states].

Recall that in this section of his paper Tononi argues for equating consciousness with integrated information roughly because the perspectival and holistic nature of consciousness mirrors the holistic and "perspective-generating" properties of the II polytrope in Q-space. It is in virtue of the (holistic and perspectival) properties of qualia modeled in II terms that the account becomes about consciousness at all, which I’ve made explicit in the brackets above in 1. Again, Tononi’s position is that an account of qualia is an account of consciousness. Thus 1 above boils down to:

2. consciousness is nothing more than each (conscious) experience having a q-character.

But this (as detailed in section 3.1) collapses *a priori* the distinction between SIL-consciousness (s-character) and phenomenality/qualia (q-character); it is a departure from our pre-theoretic understanding of consciousness and qualia that isn’t warranted by the empirical considerations offered by Tononi. That said, consciousness (s-character) might turn out *empirically* to be tied to the qualia (q-character) in that an experience’s q-character might turn out to have a common underlying explanation that makes it coextensive with s-character. And further, s-character might turn out to be independent from q-character, especially if q-character is II and s-character is (say) along the lines of a
global workspace account. Even having addressed qualia in II terms, one has considerably more work remaining in order to extend the account to consciousness, and it ought to be empirical work, not merely the “defining in” of the solution that seems to tempt many.

How might the ongoing confirmation of an IITQ lead to an account of consciousness? Presumably IITQ would be further confirmed by further mapping the integrated connections of the brain, by formally characterizing the Q-spaces associated with these neural complexes, and by using the formal Q-space models to predict/retrodict qualitative aspects of consciousness. This process of confirmation could well lead to “collateral” discoveries about the s-character of experience, especially since the qualitative and subjective character of our experience seem so intimately connected. We might discover that the Q-space models of the qualia leave nothing further to be explained by broader SIL-concepts of consciousness; this might sound like a reduction, but it isn’t a reduction to *qualia*, rather it would be something more like an elimination, but the point to note is that it would be something other than an *a priori* defining away (or defining in) of SIL-consciousness. Alternatively, the fully confirmed II account and Q-space models might lead to the discovery of higher-level (irreducible) II dynamics that could be identified with consciousness. Or (as I suggested above in section 3.2) we may learn that something like a global or centralized or highest-level feature of brain organization is required to supplement the II account with respect to consciousness. In any event, these possibilities need not impugn the present II account of qualia.

**4. Prospects: Is Integrated Information Fit to be Fundamental?**
I now examine the sense in which II is a “fundamental” and “intrinsic” property and its prospects for empirically contributing to the consciousness question. Tononi (2008) discusses explicitly the “fundamentality” of consciousness in a section of his paper (232-233), but here it seems to boil down to the idea that if one takes II as consciousness, then since consciousness exists, it follows that “an equally valid view of the universe” is to see it in terms of II, instead of mass, for example. In fact this may be a more valid view (according to Tononi), “since to be highly consciousness (to have high Φ) implies there is something it is like to be you, whereas if you just have high mass, charge or energy, there may be little to nothing it is like to be you,” and hence such entities “exist in a stronger sense than entities of high mass” (233). Insight as to how Tononi understands II’s fundamentality can be gleamed from his discussion of II’s intrinsic nature. There, Tononi writes, “Consciousness, as a fundamental property, is also an intrinsic property” (233). The idea of course is that “there being something it is like...” is paradigmatically (though contentiously) an intrinsic property, so if II is to be consciousness/qualia, it ought to be intrinsic too. But this is a rather specialized sense of intrinsic Tononi is utilizing here because as he puts it, this “means that a complex generating integrated information is conscious in a certain way regardless of any extrinsic perspective” (233). Putting these two ideas together—it seems that Tononi is really asserting only that II is a feature of a complex that exists (fundamental) and does so independently of the perspective of any observer (intrinsic), unlike un-integrated information. Thus the key assertion here is that II is an objective property—an observer independent (real) property.

Again, I am not concerned here with the philosophical debate about what counts as a fundamental (or intrinsic) property; instead I’m out to understand Tononi’s sense of
fundamental and then assess how this helps/hinders his account. I suggest that we understand II more along the lines of a (higher-level) property such asfitness (in evolutionary biology) than a property like mass. This makes sense because within Tononi’s account II doesn’t “behave” like properties usually thought of as fundamental. How much II is present depends upon the level under consideration—it can be found only at appropriately high levels where there are complexes acting as “mechanisms.”  

Additionally, in contrast to the mass of a composite, which is reducible to the mass of its components, II and fitness are not “reducible” all the way down: the complex (organism) has the property of II (fitness) and its constituent parts do not. This is the behavior of higher-level, functional, organizational, relational, or even “emergent” properties. Thus, in a strict philosophical sense, like fitness, II is relational (not intrinsic) and higher-level (not fundamental), though still an objective property.

But even if II is more like non-fundamental properties like fitness, there is still an important sense in which such a property is fundamental: fitness does play afundamental role in population genetics and evolutionary biology—that is, these theories accord a fundamental explanatory place to the notion of fitness. So while fitness may be such that an organism’s fitness supervenes on lower-level (more “fundamental”) properties, nevertheless, for explanatory purposes (within the particular theory), fitness is “fundamental” in the sense of being theoretically indispensable or irreducible. If this were the sense in which Tononi understood II as fundamental (as he well may), could II do the theoretical work Tononi requires of it? That is, what if II is onlyexplanatorily fundamental and supervenes on lower-level (more fundamental) properties (e.g., neurophysiological)? I will set aside this question for IITC and focus instead on II as an account of qualia (IITQ).
Qualia (pre-theoretically and intuitively) resist reduction/identification to mere neurophysiological properties (as discussed above), and one of the apparent virtues of the II proposal is that it doesn’t reduce qualia (directly) to neurophysical properties. It finesses the reduction by equating qualia with an informational property, not a purely physical one. For this to work, informational properties have to be fundamental, at least in the sense of not being identical/reducible to physical properties. But informational properties are quite likely to be supervenient on the physical properties of the system under consideration, after all it is in virtue of the current state’s ability to “redistribute” the probability distribution of the previous state via the causal properties of the mechanism that it carries information at all. Thus the II account would appear amenable to physicalism in one sense or another, though it does introduce an unusual feature from a physicalist perspective: a lack of definiteness.

Despite protestations to the contrary, Tononi’s conception of II seems to be relative to the spatiotemporal grid one imposes to measure II. Tononi concedes that the working hypothesis must be that “integrated information is generated at multiple spatiotemporal scales,” (235) though “there will often be a privileged spatiotemporal ‘grain size’ at which a given system forms of highest $\Phi$—the spatiotemporal scale at which it [the system] ‘exists’ most in terms of integrated information, and therefore of consciousness” (236). He goes on to suggest that preliminary understanding of the brain would suggest that the optimal spatial level for high $\Phi$ is the “grain size of neurons or minicolumns,” and similarly at the temporal level, around “tens to hundreds of milliseconds, the firing pattern of a large complex of neurons should be maximally predictive of its previous state” (236).
This is importantly different from the higher-level irreducibility discussed above. It isn’t just that there is a level at which it doesn’t make sense to ask about II, but rather that one obtains different (nonzero) answers to how much II is present at different levels/grain sizes. This is clearly not the case with mass, nor is true of fitness in the same sense.\textsuperscript{29} This lack of definiteness with respect to II would seriously undercut it as the key to consciousness. This is because some other consideration (e.g., the highest or the most central, etc.) has to be brought in to justify that a given level is the one whose II corresponds to consciousness, and thus opens the door for some other property to be at least jointly determinative of the degree of consciousness.

But these problems are largely obviated when the theory is “reigned in” to the more modest IITQ. First, the IITQ need not embrace fundamentality/intrinsicness to the same extent. In fact there are plenty of philosophical proposals that back away from qualia as intrinsic.\textsuperscript{30} The irreducible, explanatorily fundamental, and objectively real conception of II developed by Tononi may not “fit the bill” for consciousness as it is generally construed, but it does provide a philosophically defensible and potentially ground breaking empirical framework for investigating the qualitative properties of experience (qualia).

Second, multiple-level indefiniteness is not necessarily a stumbling block for qualia because the quantity of qualia isn’t really the issue, the structure of it is. It is not even clear what the quantity of qualia would mean. Qualia in conscious experience is naturally dependent on an s-character, so the details of subjectivity—both phenomenologically and ultimately neurologically—will serve to pick out the spatiotemporal grain size/level of II that is germane to the particular qualia questions under consideration, e.g., the color of the red, the feel of the tickle, the richness of the note. Additionally, as Tononi notes, qualitative
experiential “fields” are often complex, not qualia in the “strict philosophical sense,” but rather composites of such qualia (230-31), so the fact that sub-complexes and sub-sub complexes (etc.) each have “nesting” Φ-values is merely a sign of this.

Thus the more modest IITQ program is promising empirically and if ultimately confirmed may help dissolve the philosophical puzzle of qualia by “peeling it off” from the “harder” problem of consciousness, much as memory or language have been “peeled away” from it. The IITQ locates the qualitative properties of experience in the brain (like memory) and yet still intimately tied to consciousness (also like memory), via its s-character. It is an important and novel feature of Tononi’s work that in “peeling off” qualia from the problem of consciousness, he is able to do so without definitionally excising its (potentially) deep connection with consciousness—a connection that may well turn out to be an important step to understanding consciousness itself.31

A further interesting feature of qualia on the II account is that they aren’t precisely functional properties in the standard sense, but rather as the formal informational properties of Q-space are what might be called “implementational properties.” As discussed above, Tononi understands these properties themselves as supervening on the physical brain/body system and hence “allows for” qualia to be investigated empirically by science. Philosophers often mistake the II approach for a functionalist account since it doesn’t analyze qualia/consciousness in messy neuroscience terms. But a careful read of Tononi makes clear that two functionally equivalent neural arrangements can differ in their internal integrated information and hence won’t have the same quality of experience. This could happen if the functional neural units had different levels of redundancies or if they implemented the function using a distinct but equivalent logical form (a negated
disjunction of negated or-gates instead of an and-gate). The messy details of the
implementation of the system do matter, if not to behavior then certainly to consciousness.

Precisely how these implementation properties are related to the functional
properties of the brain is certain to be rather complicated. Presumably they are in one
sense at a lower level than functional properties—along the lines of the hardware/software
dichotomy. So for example, the functional units in the visual system are implemented by
various neuronal complexes with their own effectively unique implementational
(II/qualitative) properties. But as William Lycan (1987) has made clear, such “levels” are
not likely to be discrete or monotonic; there will be mixed level functional (and
implementational) complexes that interact with levels both “above” and “below” and
“levels” may well turn back on themselves. Developments in neurodynamics have further
confirmed the importance of understanding the brain on multiple, involuted, and even
indistinct levels.32

5. Closing Thought
The IITQ approach identifies qualia with the integrated informational properties of the
brain. The qualitative properties of conscious experience are understood as infomational,
not functional. On this understanding the IITQ account is not directly about conscious
experience (SIL-consciousness); rather it offers an account of the various qualitative
aspects of conscious experience. This is an important problem in its own right of course,
and it may even turn out to be the key to consciousness, but so construed the IITQ account
doesn’t conflate qualia and consciousness, nor does it assume that an account of qualia
immediately yields an account of consciousness. Depending on the confirmational details
of IITQ and our emerging understanding of information and brain dynamics, it may turn out that the key to consciousness is emergent dynamical properties, global information integration, or something else yet to be discovered. Alternatively, things may go a more deflationary way, namely that further work on IITQ and the various functional (autobiographical memory, volitional, emotional, etc.) capacities of the brain reveal that the predicate “SIL-consciousness” doesn’t have a referent once qualia and the functional capacities are understood. In either case the II account has opened up a way to decouple the fates of qualia and consciousness.

6. References


Peressini, A. (forthcoming) “Dis-Integrating Intuitions about Consciousness and Qualia.”


* I thank Melanie Boly, Tony Hudetz, Jodi Melamed, and Giulio Tononi for their various contributions to this paper. I am also grateful for the comments of anonymous referees and my philosophy research seminar students at Marquette University in the fall of 2010.
In applying information theory, Tononi (2008) sets it up so that the informational focus is always backwards in time. Given a mechanism and current state, “the system’s mechanism and state constitute information (about the system’s previous state), in the classic sense of reduction of uncertainty or ignorance” (220). This decision seems both substantive and arbitrary, with respect to consciousness. Why not frame the application of information theory in a forward direction, so the mechanism and current state give information about the system’s next state? Perhaps this has something to do with how information theory is typically developed, but given that a finite state mechanism is assumed, it seems “forward probabilities” would work as well. I return to this question below.

See Herstein (1975, 40ff) for more detail.

I presume another way of stating this is that MIP is the decomposition into parts that maximize II for the parts considered independently. It isn’t entirely clear how to understand MIP because it is characterized as both the partition that leaves the least information unaccounted for and the partition of the system into minimal parts; it is not hard to imagine that these might be equivalent, but something ought to be said to establish this in general.

Making use of the actual “causal mechanism” to ground eligible partitions raises its own questions, the first being, why think the causal approach is univocal? It may well be that given the complexity (on multiple levels) of the brain/body system, no unique causal schematization exists either. And what is more pressing, if neurodynamicists like Walter Freeman (2001) are on the right track, then the entire notion of “causal mechanism,” traditionally construed has no application to the brain/body system. See Freeman (2001, esp. 126-140) for discussion.

Actually the partition need not be unique as long as it can be established that each of the non-unique maximums yield the same entropy relative to $p(X_0(mech, x_1))$, but again this is a far from trivial result to establish. The possibility of this lack of uniqueness is acknowledged in Balduzzi and Tononi (2008, 7) and there they stipulate without explanation that the partition with the lowest (un-normalized) effective information be taken to be the MIP.

I return to the question of “grain size” below in Section 4. I am grateful to an anonymous reviewer for this journal for this insight regarding Tononi’s possible response.

Barrett and Seth (2011) have recently generalized a version of $\Phi$ for less idealized continuous systems but in doing so back off both the claim that their version of $\Phi$ can be interpreted as
identical to consciousness (13-14) and even that it can be given a physical interpretation at all. (15).

The problem that arises for infinite cases is that the interpretation of the probability calculus becomes trickier, since further formal constraints (with problematic scientific interpretations) become necessary to not violate the formal calculus and retain finite maximum entropy distributions. See Sec. 2.2 below for discussion of interpretive difficulties with probability.

There is reason to think such exponential increase in dimension can both be physically meaningful and tractable from a mathematical standpoint, since the application of knot theory in the theoretical physics formalism of M-theory has accomplished something similar. While the details are beyond the scope of this paper, these developments in knot theory may also provide a principled means of determining a definite probability distribution and thus obviating another of the formal concerns with the II account. I owe thanks to an anonymous reviewer for this insight.

Beaton (forthcoming) “Integrated Information and Internalism.”

See Eells (1991, esp. 34-55) for a philosophical discussion of interpretations of probability and further references.

As an anonymous reviewer rightly points out, to the extent that the II account is equivalent (or dependent on) understanding consciousness in terms of probability, discussions of probability in this context are closely related to discussions of meaning. A real virtue of the II approach is its natural connection to the notion of meaning and semantics (Tononi 2008, 338-9), thus Tononi’s account of meaning may well provide the key to the appropriate interpretation of probability. I limit my critique here (and revised philosophical understanding) to qualia/consciousness, but my proposal retains this natural connection to meaning. It is worth pointing out that Tononi’s discussion of meaning appears to point to a subjectivist and externalist interpretation, which seems to be at odds with his internalist operationalization of II in terms of $\Phi$. See Beaton (forthcoming) for a careful critique of Tononi’s operationalization of II.

Compare this definition to Boly and Massimini and Tononi’s (2009, 386) characterization of it as “what disappears when you go into a dreamless sleep.” I believe them equivalent for my purposes here, especially since both use an intuitive definite description that remains “topically neutral” as to the neurophysiological basis of the phenomenon. One possible difference is that
red qualia disappear when one goes into a dreamless sleep, but it isn’t obvious that red qualia are a necessary part of being conscious, so I favor the Nagel version used above.

14 The kind of properties I have in mind by “natural” are fundamental or simple properties in the rough sense in which we do believe that there is property in the world that corresponds to “being a hydrogen atom” or “being a water molecule,” but not to “being a mousetrap.”

15 For prominent examples see Block (1980), Jackson (1982), and Chalmers (1999).

16 Tim Bayne (2007) makes related points (using Searle’s terminology) between a unified field vs. building block approach to the structure of consciousness.

17 For example, the same neural mechanism generates different shapes in Q-space depending on its other connections just as the quality of experience depends on “context” (228), the modalities of experience are captured by densely entangled q-arrows in Q-space (227, 230), “elementary experiences” (simple qualia) correspond to densely entangled complexes without further entangled sub-complexes (230), overlapping hierarchies of entanglement within nested complexes correspond to overlapping hierarchies within phenomenological fields (231), the difference between “categorical” sense modalities (smell) and “topographical” ones (vision) are captured by difference basic sub-shapes in Q-space (231). See Tononi (2008, esp. 230-232) for details and further examples.

18 I am not convinced there really is a “hard problem” in Chalmers’ sense. Nonetheless the problem of consciousness concerns both the SIL-consciousness of creatures and the phenomenality of their mental states and if the II account is on the right track, then SIL-consciousness may well be the “harder” of the two.

19 Kriegel parses things a bit differently. Kriegel uses p-consciousness on the creature level to indicate consciousness taken as a whole in its full sense, much like my SIL-consciousness. But Kriegel also uses p-consciousness as a property of a mental state: a p-conscious state has an s-character and q-character, whereas on my approach, a SIL-conscious entity may have aspects of its consciousness with s- and q-characters. In either case, it is the q-character that corresponds to qualia and the s-character that presupposes SIL-consciousness experiencer.

20 There is preliminary empirical reason to doubt that on an intuitive level SIL-consciousness “entails” qualia. See Sytsma and Machery (2010) and Peressini (forthcoming) on intuitions and pre-theoretic rationales regarding qualia and SIL-consciousness.
It would also seem to follow from this reasoning that in cases of a severed corpus callosum (so called “split-brain”) there are in fact two consciousnesses. Tononi (2008, 219) “bites the bullet” and writes that in such cases “the surgery has created two separate consciousnesses instead of one.” This seems to be a departure from reported first person experience in such cases and also from the how we tend to think of SIL-consciousness from a pre-theoretical perspective.

As Ned Block (2009) argues, the explanatory gap should be present in any responsible account at this stage since science hasn’t yet provided us with the same sort of enriched conceptual apparatus that allowed us to understand heat as motion and light as vibration in ways the ancients couldn’t imagine. As I will argue below, IITQ may be an initial step in that direction.

It should be pointed out that Tononi’s empirical and phenomenological evidence could well be consistent with the discovery that SIL-consciousness is (for example) some sort of global workspace, which in the case of creatures like us is generated out of highly integrated informational neural complexes, hence the correlation between high II and SIL-consciousness.

Shaun Nichols (2011) has a different but still generally supportive (of my point above) account of our resistance to mind/brain identity claims that focuses on how our “low path” neuro/cognitive wiring that intuitively recognizes conscious beings won’t likely ever be engaged by the theoretical reduction of consciousness to brain activity. If he is right then even a finely grained complex reduction is going to leave us feeling at least partially unsatisfied with a scientific account.

Tononi gives up a standard understanding of qualia’s intrinsicness when he asserts that the univocal specification of qualia by the Q-space shape shows that “the ‘inverted spectrum’ is impossible…” (241, note 8).

Returning again to the question about the direction of information “flow” (note 1), the choice of focusing on the previous state in defining II does seem to make II dependent on the external perspective of whether one considers it to be information about the previous or subsequent state, and thereby undercut the claim of intrinsicness.

As I noted in Section 2.1, it is a bit tricky reading Tononi on this point because in places he seems to assume there is a physically privileged and fundamental mechanistic description of every physical system. If so, then some qualification in my indeterminacy critique may be necessary. Nonetheless, even granting such a fundamental level for the privileged mechanistic
description, my point that the components of this mechanism themselves do not have II (unlike mass) goes through.

28 In general, Tononi does intend informational properties to supervene on the physical, but he does make one reference to the idea that information may be “in an ontological sense, prior to conventional physical properties . . .” (233). By this he intends a informational interpretation of quantum physics and cites Wheeler and Ford (1998).

29 It is true that an organism has a fitness and so do some of its proper parts, alleles, for example, but this is rather different since fitness, properly speaking, applies fundamentally to individual organisms and only derivatively to alleles, i.e., the fitness of an allele is defined as the average fitness of the organisms possessing it. Even if one is a pluralist about the level(s) of selection and understands genes as having fitness nonderivatively, there is still a level below genes the elements of which do not have the property of fitness.

30 See for example Johnsen (1997) and Nikolinakos (1994) and Tye (2007), and of course Dennett (1988). It would be an interesting project to work out the extent to which this interpretation of Tononi’s IITQ was anticipated by Dennett’s (1988) phenomenological information property (PIPs) detectors, which were in turn inspired by Dretske’s information (1981) account.

31 It may also turn out that after “peeling away” qualia, nothing much is left of the harder problem of SIL-consciousness, being open to this as an empirical possibility is yet another positive feature of the II approach.

32 See for example Walter Freeman’s (2000) work on neurodynamics with its interacting micro, meso, and macro levels.