Contextuality in the Integrated Information Theory

J. Acacio de Barros[1,3], Carlos Montemayor[2], and Leonardo P. G. De Assis[3]

[1] School of Humanities and Liberal Studies, San Francisco State University, San Francisco, CA
barros@sfsu.edu
[2] Department of Philosophy, San Francisco State University, San Francisco, CA
cmontema@sfsu.edu
[3] Suppes Brain Lab, Center for the Study of Language and Information, Stanford University, Stanford, CA
lpgassis@stanford.edu

Abstract. Integrated Information Theory (IIT) is one of the most influential theories of consciousness, mainly due to its claim of mathematically formalizing consciousness in a measurable way. However, the theory, as it is formulated, does not account for contextual observations that are crucial for understanding consciousness. Here we put forth three possible difficulties for its current version, which could be interpreted as a trilemma. Either consciousness is contextual or not. If contextual, either IIT needs revisions to its axioms to include contextuality, or it is inconsistent. If consciousness is not contextual, then IIT faces an empirical challenge. Therefore, we argue that IIT in its current version is inadequate.

Key words: consciousness, contextuality, integrated information theory

1 Introduction

The Integrated Information Theory (IIT), developed by Giulio Tononi in a series of influential papers, promises to deliver not only an account of consciousness but also a concrete way to measure it [20,25]. In this paper, we focus on the second aspect of IIT. We shall argue that there are potential problems that require clarification concerning a tension between IIT's mathematical model of consciousness and contemporary models of contextuality.

Contextuality is important to measure consciousness for several reasons. Here we mention three salient ones. As IIT makes clear, the integration of semantic content is fundamental for understanding consciousness. IIT is so explicit about this semantic integration that it provides a set of definitions regarding not only conceptual content, but also how conceptual content is integrated into maximally specific experiences. But conscious content is always determined at a context of informational background, which is cognitive and perceptual, with many variables that need to be determined at any moment in time. This is one of the main

reasons why contextuality is central in linguistics and pragmatics—content depends on context and background assumptions.

A second reason to assume contextuality as a constraint on theories (and more pressingly measurements) of consciousness concerns attention. Attention, like conversational content, depends on background conditions and relevance. It also depends on conceptual content and is guided by many neural processes associated with voluntary and involuntary attention [19]. Context, therefore, is not only environmentally driven, but also motivationally determined.

Finally, a very important reason to take contextuality seriously into account in a theory of consciousness is the very nature of measurements. Measurements are notoriously contextual, a fact made quite vivid not only by quantum mechanics, but also by psychology and linguistics, disciplines who occupy a central role in studies about consciousness [3,17,6]. Thus, the contextuality of measurements is central to one of the main goals of IIT: to provide a measure for consciousness.

Our main argument is that IIT is empirically problematic, based on formal considerations concerning contextuality. One possibility is that IIT is problematic because, in its current formulation, it is incompatible with our current understandings of how to accommodate mathematically content that is contextual due to limited access to all processes. This would mean that IIT is in principle plausible, but in practice impossible to test. Alternatively, it could be that IIT is incomplete and needs critical amendments. This would mean that the theory could be compatible with our current understanding of contextuality but that it is unclear how it could be compatible with it. In either case, we believe that IIT is empirically problematic as it stands now, and that clarification is needed.

A more troublesome possibility, for which we will not argue as decisively as the previous empirical one, is that, in its current version, IIT is in principle incompatible with mathematical approaches to contextuality. Given the centrality of contextuality in understanding consciousness, this would make IIT internally inconsistent. This would mean that IIT needs to be abandoned. We will not develop this criticism and will only focus on the empirical one, but we mention this problem because we believe this is also an issue that demands further clarity, namely, it needs to be demonstrated that if IIT turns out to be incomplete, that it is at least in principle compatible with contextual data.

Before proceeding, two crucial clarifications are needed. First, our criticism is based on considerations concerning a notion of contextuality susceptible to mathematical analysis. IIT explicitly demands a mathematical treatment of the Φ measure and, as we will explain, this demand entails a mathematical treatment of contextuality. Our criticism only targets this formal, but still crucial, aspect of IIT. Because of our focus, whatever metaphysical commitments IIT has, for instance regarding panpsychism (or dualistic and monistic interpretations), are beyond the scope of this paper. Second, the notion of contextuality we work with here is relevant to linguistics, but we are not appealing to all cases of context sensitivity in linguistics. Rather, we use only a restricted sense of contextuality that can be formalized in terms of violations to sums of probabilities. Thus, we do

not address forms of context dependence in pragmatics and forms of implicature in general.

To put forth our argument, we organize this paper in the following way. First, in Section 2 we discuss the current understanding of the mathematical theory of contextuality. Then, in Section 3, we present discussions of contextuality in IIT, and put forth our main argument. We end with some comments and discussions in Section 4.

2 Contextuality

Contextuality is an important concept in many different fields, such as in linguistics, physics and psychology. For that reason, there are many different definitions of contextuality, but here we focus on a mathematically precise definition that is relevant to IIT, as it directly relates to theories of measurement. Intuitively, contextuality is the idea that a quantity (say, the truth value of a proposition) depends on the overall environment in which it is present. To formalize such idea, the concept of random variables is used.

First, let us start with probabilities. The most straightforward way to define probabilities is axiomatic [18]. Accordingly, a probability space $(\Omega, \mathcal{F}, p)$, where Ω is a a set of possible elementary events (the sample space), $\mathcal{F}$ an algebra (of events) over Ω, and p a function $p : \mathcal{F} \to [0, 1]$, is a triple satisfying:

1. $p(\Omega) = 1$
2. $p(\bigcup_i A_i) = \sum_i A_i$, for $i \neq j$ and $A_i \cap A_j = \emptyset$.

In this definition, $p\,(A)$ is the probability of event A happening.

A random variable $\mathbf{R}$ is a (measurable) function $\mathbf{R} : \mathcal{F} \to E$, where E is a set of real numbers. The idea of a random variable is to model the stochastic properties of the outcomes of a given experiment, where $e \in E$ corresponds to possible values of such outcomes. For example, imagine a hypothetical experiment measuring participants heights, with the minimum measurable height being 110 cm and the maximum 210 cm, with a resolution of 1 cm, the set of possible outcomes of measurements is $E = \{110, 111, 112, \ldots, 209, 210\}$. If we randomly select participants, the outcomes of their height measurement will follow some distribution (perhaps two superposed and truncated Gaussians, corresponding to female and male participants). A random variable modeling the height-measuring experiment should have all the same stochastic characteristics of it, and the random sampling of elements of Ω corresponds, intuitively, to the random sampling of participants and their respective heights.

The range of values of a random variable can be set to match that of any type of measurement, but the simplest ones are yes-no questions (e.g., "is this person taller than 170 cm?"). For such cases, two-valued random variables may be used to correspond to answers to the question "does the object/system have property P?". For example, E may be chosen as being either -1 (for "no") or 1 (for "yes"). If the the property P is measured, then we record 1, and if it does not, we record -1. Since this is modeled with the random variable $\mathbf{R} : F \to \{-1, 1\}$ on a probability

space $(\Omega, \mathcal{F}, p)$, we can think about the two-valued random variables as truth-values for propositions about the system, and the algebra $\mathcal{F}$ as corresponding to logical statements about such propositions (e.g. for two distinct elements $A_1, A_2 \in \mathcal{F}$, $A_1 \cup A_2$ and $A_1 \cap A_2$ are also in $\mathcal{F}$, and correspond to the logical connectives "or" and "and," respectively). In other words, random variables (and their corresponding probabilistic measures) correspond to a natural (stochastic) extension of logical statements about the nature of experimental outcomes. The logical structure of the statements come from the underlying Boolean algebraic structure that is derived from the ordering provided by the probability function p over the algebra $\mathcal{F}$.

So, how does contextuality come about in the language of random variables? As mentioned above, a system is contextual if it varies from one context to another. But what do we mean by "vary," and what do we mean by "context?" Let us start with an example, which will be useful below. Imagine we have a set of N properties (or concepts), denoted by P_i, $i = 1, \ldots, N$. The simplest contextual example could be though of as coming from $N = 2$, as in what happens with order effects. For example, consider the following two questions reflecting participants beliefs, discussed by [26,27]: P_1 = "Do you generally think Bill Clinton is honest and trustworthy?"; P_2 = "Do you generally think Al Gore is honest and trustworthy?" Since those are two separate questions, they must be asked sequentially. We have only two possible ways to ask those questions: first P_1 and then P_2 or first P_2 and then P_1. It so happens that when doing so, the probabilities for P_i change. For instance, in a 1997 Gallup pool [26,27], respondents answered yes to P_1 at a rate of 50% when P_1 was first, and 57% when P_1 was asked after P_2. Similarly, P_2 got a rate of 68% when first, and 60% when after P_1. This clearly shows an order effect, but more importantly, in a certain sense Clinton was considered by respondents as more trustworthy in the *context* of his relation with Al Gore than not, whereas Gore lost some of his trustworthiness when associated to Clinton.

In terms of random variables, if we think of $\mathbf{P}_1$ and $\mathbf{P}_2$ as representing those questions, then we have changes in the expectations of those random variables according to their order (or context). This is a situation where we have direct influences of one variable (which may also establish context) onto another. For example, in our Clinton/Gore example, we can think of the question P_1 (or P_2) directly influencing the respondent's belief about the following question: Gore gives Clinton a honesty bump. We call this explicit contextuality[4].

A more subtle case occurs when the random variables are not inconsistently connected. To see this, let us examine $N = 3$, and also that the properties are "yes" or "no." This is described by ± 1-valued random variables, $\mathbf{P}_1$, $\mathbf{P}_2$, and $\mathbf{P}_3$, whose expectations are all equal to zero, meaning that we have equally random chances to either get $+1$ or -1 as outcomes of measurements of those variables. This case is more interesting because, as constructed, we do not allow for the type of explicit contextuality discussed in the paragraph above. Vari-

[4] This term was introduced by Pawel Kurzynski. See his contribution to this conference.

ables $\mathbf{P}_1$, $\mathbf{P}_2$, and $\mathbf{P}_3$ may be correlated: their pairwise joint expectations (e.g. $E(\mathbf{P}_1\mathbf{P}_2)$) can take values between -1 and 1, corresponding to anti-correlated and perfectly correlated (with 0 meaning that they have no correlation)[5]. Imagine furthermore that experimental conditions are such that we can never observe all three random variables together, but only in pairs. It is possible to imagine an experimental setup that the measured correlations, given the impossibility of simultaneous observations of all three variables, be, for example, $E(\mathbf{P}_1\mathbf{P}_2) = E(\mathbf{P}_1\mathbf{P}_3) = E(\mathbf{P}_3\mathbf{P}_2) = -1$ (for a concrete example, see [5]). It is easy to see that there is a problem with the -1 correlations. For example, if $\mathbf{P}_1 = 1$, the first correlation implies $\mathbf{P}_2 = -1$, and the third implies that $\mathbf{P}_3 = 1$, which in turn, from the second correlation, implies $\mathbf{P}_1 = -1$, a clear contradiction. What is leading to the contradiction is the assumption that the variable $\mathbf{P}_1$ in the context of the experiment measuring $(\mathbf{P}_1, \mathbf{P}_2)$ is the same as the $\mathbf{P}_1$ in the context $(\mathbf{P}_1, \mathbf{P}_2)$[6]. If we were to, for example, index the variables (as proposed by Dzhafarov and Kujala [9,10][7]) according to their context, such contradictions would not appear.[8]

The above example shows how contextuality might be manifest as the impossibility of assigning the same values to a quantity in a way that is independent of the context. However, as it is presented, it comes from a logical contradiction. So, the question remains as to how one can extend the criteria for stochastic systems. A way to see this comes from the work of Abramsky and Hardy, where they showed that violations of logical consistency such as the one above are necessary and sufficient conditions the non-existence of a joint probability distribution (jpd) [1]. In other words, even when we have probabilistic outcomes, the existence of a single probability space $(\Omega, \mathcal{F}, p)$ is a necessary and sufficient condition for no logical inconsistencies, and therefore no contextuality. As a consequence, for the example of three variables, it can be shown [24] that the variables are not contextual iff

$$\begin{aligned} -1 \leq\ & E(\mathbf{P}_1\mathbf{P}_2) + E(\mathbf{P}_1\mathbf{P}_3) + E(\mathbf{P}_2\mathbf{P}_3) \\ & \leq 1 + 2\min\left\{E(\mathbf{P}_1\mathbf{P}_2), E(\mathbf{P}_1\mathbf{P}_3), E(\mathbf{P}_2\mathbf{P}_3)\right\}. \end{aligned} \tag{1}$$

The logical violation is a more subtle example of contextuality than the first one examined, where the statistical properties of a quantity changed with context. To distinguish the two types of contextuality above, one that is manifest

[5] Because of our choice of ± 1-valued random variables with zero expectation, their joint expectations coincide with their correlations.

[6] This example is examined in detail in Specker's Parable of the Over-Zealous Seer [cite], but was also discussed much earlier on by Boole [cite].

[7] The indexing idea is also related to Stalnaker's two-dimensional semantics; see [23].

[8] In the works of Dzhafarov and Kujala, when we can assign contextuality because of direct influences between the measuring conditions of random variables, such variables are said to be inconsistently connected [10,14,8,12,11]. To those author's, a system is contextual only if all context effects are not explainable by direct influences. So, for them the $\mathbf{P}_1$, $\mathbf{P}_2$, and $\mathbf{P}_3$ perfectly anti-correlated example is contextual, whereas the Clinton/Gore one is not. However, we emphasize that this is a nomenclature issue.

in the changed expectations from one context to another, and the other that is a consequence of the impossibility of attaching a consistent underlying logical structure via a jpd, we refer to systems that exhibit the former as exhibiting *explicit contextuality* or being *explicitly contextual* (or, according to [13], *inconsistently connected systems*) and the latter as exhibiting *hidden contextuality* or being *implicitly contextual*.

The example above can be generalized to more than three random variables, as well as to random variables that take multiple values. To represent this in terms of random variables in a way that makes the context explicit, we can use a contextual index in the following way. We start with the assumption that each experiment and its corresponding variables correspond to a context. We think of the random variables as contextual when we cannot associated to a variable P_i in one context the same probability space as the P_i in another context (i.e., there is no single jpd that describe P_i in all contexts). For our three random variable example above, only pairs are observable, namely (P_1, P_2), (P_1, P_3), or (P_2, P_3), but never triples, e.g. (P_1, P_2, P_3). Let us call C_1 the experimental condition (or context) (P_1, P_2), C_2 condition (P_1, P_3), and C_3 condition (P_2, P_3). To represent this explicitly in our notation, we add an index for context. For example, $P_{1,1}$ is P_1 in context C_1, whereas $P_{1,2}$ is P_1 in context C_2, and so on.

With this notation in mind, inconsistently connected systems are those in which it is not true that $\mathbf{P}_{1,1} \sim \mathbf{P}_{1,2}$, where this notation means "the random variable $\mathbf{P}_{1,1}$ has the same distribution as $\mathbf{P}_{1,2}$." As an example, let us revisit the Cliton-Gore order-effect survey, where two questions are asked in sequence in two different orders, C_1 and C_2:

C_1**:** $P_{1,1}$ = "Is Bill Clinton trustworthy?"; $P_{2,1}$ = "Is Al Gore trustworthy?";
C_2**:** $P_{2,2}$ = "Is Al Gore trustworthy?"; $P_{1,2}$ = "Is Bill Clinton trustworthy?".

It may be somewhat surprising that the expected answer to $P_{1,1}$, denoted by $E(\mathbf{P}_{1,1})$ and given by

$$\begin{aligned} E(\mathbf{P}_{1,1}) &= \sum_{\omega_i \in \Omega} p\left(\omega_i\right) \mathbf{P}_{1,1}\left(\omega_i\right) \\ &= p\left(\mathbf{P}_{1,1} = 1\right) - p\left(\mathbf{P}_{1,1} = -1\right), \end{aligned}$$

is more positive toward Bill Clinton than $P_{1,2}$, but that is what was shown empirically [22] (i.e., $E(\mathbf{P}_{1,1}) > E(\mathbf{P}_{1,2})$). However, we should point out that mathematically, because we are using contextual indexing, it is not problematic to have different expectations for each context, whereas in the example with no contextual indexing, we would have a seemingly direct contradiction ($\mathbf{P}_1 \nsim \mathbf{P}_1$).

In the hidden contextuality case with three random variables, the indexed notation can also be extended. As before, imagine the extreme case where $E(\mathbf{P}_{1,1}\mathbf{P}_{2,1}) = E(\mathbf{P}_{1,2}\mathbf{P}_{3,2}) = E(\mathbf{P}_{2,3}\mathbf{P}_{3,3}) = -1$. If we do not assume that the observed property is independent of context, we run into no problems. However, if we set $\mathbf{P}_{i,j} = \mathbf{P}_{i,j'}$, we run into the same type of problems as before, and reach a contradiction.

At this point it is worthwhile to discuss some aspects of contextuality that are directly relevant to our argument. Contextuality only exists when we cannot observe all quantities of interest simultaneously: measuring all random variables at the same time implies the existence of a jpd by simply creating a data table that can be used to compute the jpd and the relative frequencies for each marginal distribution. However, it is often the case that the random variables cannot be all measured simultaneously. This lack of simultaneous measurement may have two different origins: (i) it may be impossible *in principle* to measure $\mathbf{P}_1, \mathbf{P}_2, \ldots, \mathbf{P}_N$ simultaneously, or (ii) it may be empirically difficult, perhaps even impossible in practice.

For (i), there can be situations, particularly when the system is contextual by direct influences, where the observation of $\mathbf{P}_1$ precedes temporally and affects directly $\mathbf{P}_2$; this seems to be the case for the example of the Gore/Clinton questionnaire discussed above. One cannot ask a question about Al Gore's trustworthiness simultaneously with a question about Clinton; they must be asked in order. The same is the case for contextual examples in quantum mechanics, where there is no jpd. In entangled quantum systems, there are no direct influences, but we cannot measure non-commuting observables simultaneously, and depending on the choice of observables, no jpd exists [15].

For (ii), the situation is slightly different. There may not be a principled reason for not observing all three random variables simultaneously, but experimental design or measurement constraints may create a *de facto* impossibility of observing them. This was, for instance, the case of the contextual firefly introduced by Foulis (see [5] for an explicit contextual model). Because experimental constraints or technical limitations prevent us from observing all variables at the same time, the correlations between them may be enhanced, such that by this process the observations cannot fit a jpd. In other words, the marginal expectations of correlations change from context to context, in a way similar to explicit contextuality. Furthermore, it might be possible that when we observe a system, we may be unaware of the random variables being contextual, which is something that we only verify empirically. For instance, there is nothing strange about observing correlations $E(\mathbf{P}_1\mathbf{P}_2) = E(\mathbf{P}_1\mathbf{P}_3) = E(\mathbf{P}_3\mathbf{P}_2) = -1$. It is not until we put all three together, in an attempt to obtain a jpd, that we realize their inconsistency. That no jpd exists in certain circumstances is nontrivial for the three-random-variable example, and it becomes even more difficult to establish for more variables (consistency is checked with the satisfaction of inequalities whose complexities increase rapidly with the number of involved variables).

Contextuality shows up in many situations, from quantum mechanics to social sciences (see [17,16,6] and references therein). In particular, well-known examples exist in cognitive sciences, where human decision making has been shown to not follow classical probability theory, being therefore either explicitly or implicitly contextual (though recent work suggests that in psychology all examples are explicitly contextual [7]). Furthermore, the connection between speech and thought is well know to be contextual, with many examples discussed in the literature. Thus, a discussion of contextuality as it refers to proposed theories

of consciousness is not only relevant, but essential. In the next section, we will turn our attention to one such theory, the Integrated Information Theory (IIT) [20].

3 Contextuality in IIT

It is unclear whether the measurable human mind (thought here to be equivalent to the brain and its physical states) is contextual in principle. It is not unimaginable (in fact, many theories do so) that, for example, quantum processes are important in the brain. If this is the case, it is possible (though we believe improbable, mainly due to decoherence) that entanglement of relevant neuronal processes exist. Such entanglement may produce contextual random variables, and would preclude the existence of a jpd.

Though the previous argument could be made that the brain is contextual, at the microlevel, we want to focus on the difficulties mentioned in Section 2. First, why should we bother with contextuality, at least from an empirical point of view, when thinking about the brain. The main reason is that contextuality shows up in many situations in the social sciences (see [6] and references therein). In particular, well-known examples exist on cognitive sciences, where human decision making has been show to not follow classical probability theory, in the sense of being incompatible with a jpd, therefore being contextual. Furthermore, as mentioned above, the connection between speech and thought is contextual. In other words, behavioral outcomes are contextual, and ultimately what we observe is tied, in one way or another, with behavioral outcomes.

It is possible that such contextuality comes from factors unknown to or uncontrollable by the experimenter. For example, in a real-world situation, where most learning happens, the brain is bombarded with huge amounts of disparate stimuli, some of them perhaps even seemingly contradictory to each other (e.g., simultaneous exposure to stimuli that represent pain and pleasure). Such stimuli are not forgotten, insofar as learned unconscious decision processes are concerned, by moving to a new environment in the protected conditions of controlled experiments. If we now think in terms of brain mechanisms, the presentation of a stimulus may activate not only neurons associated with this stimulus, but also other context-relevant neurons that were activated in the learned real-world situations. Furthermore, because we should expect neural activation to be stronger to the original stimulus, the detection of such neural patterns would be very difficult (particularly because we would not know what we should be looking for). If we could, perhaps, be able to measure all neurons in the brain, we would not have any implicit contextuality showing up (at least not in the measured firing patterns), though we could have explicit contextuality; however, as we will see below, this is very difficult empirically.

To see this, let us examine the well-known case of the "guppy" effect [21,2] in concept combination. The guppy effect refers to the established fact that when participants are asked to name objects that belong to the concept "pet" and objects that belong to "fish," guppies appear with very low frequency. However,

when asked to name objects that belong to "pet-fish," guppies are high up on the list. What makes this example interesting is not that concept combination changes the frequency of "guppies," but instead that it changes such frequencies in a way that is incompatible with classical probability theory (i.e., with a jpd) [2]. In other words, concept combination as an internal process in the brain is contextual.

Now, let us say we try to approach the problem mentioned above, of measuring all the neurons associated to some cognitive process. How would we know which of the neurons are relevant. For instance, we know that once a concept (say, "fish") is presented, there is a spreading activation of neurons that are related to other concepts (e.g., "flounder," "cod," "tuna," "sushi," "Easter," etc). Such web of activated neurons is strongly present in one context, but is not in another (such as "guppy"). That means that one would have to know what to look for, at the level of neurons, even when what we are looking for is not currently active. In other words, to be able to construct a jpd, if it exists, one would have to measure everything (including external conditions that might seem irrelevant to the experimenters under the situation, such as temperature, barometric pressure, amount of saliva in subject's mouth, heart beat, etc), since any such variables could present contextual cues that are necessary for the construction of a jpd. However, as one could imagine, this would not only pose a huge measurement problem, from a practical point of view, but would also have so many variables that would make it impossible to obtain any type of statistical information about the system of interest, as every experiment would be, in a certain way, unique in terms of control variables. Of course, even if we maximally specify all those variables, it is still possible that the system displays explicit contextuality, and no jpd exists.

To summarize, we have the following empirical difficulty brought about by contextuality. Since contextuality exists in practice (numerous experiments corroborate this), we do not have a joint probability distribution. The only way to overcome the problem of contextuality would be to observe everything, clearly a daunting task. But even in such cases, however, no jpd exists, as we would move from implicit to explicit contextuality. As we will see below, those issues are a direct challenge for the current version of Tononi's IIT.

We now turn to IIT. As discussed above, Tononi's IIT is one of the most important theories of consciousness currently proposed [20]. IIT is an attempt to characterize consciousness both quantitatively and qualitatively, giving it a precise mathematical formulation. Unlike the traditional approach used in neuroscience, IIT takes as its starting point the phenomenology of consciousness, and postulates the properties the physical mechanisms, such as neurons with their synapses, shall respect so that consciousness can take place. One of goals of IIT is to quantify in what extent one system has consciousness, that is, what mechanisms belonging to that system contribute to the emergence of consciousness, and how much they contribute to it. The second goal is to build a theoretical tools able to discriminate the different kinds of consciousness that the system can display. In other words, define a qualia-space.

According IIT, a system that is capable of generating consciousness must have a high capacity to discriminate a large number of different states, which are related to the amount of information that distinct subsystems may generate. However, IIT also affirms that the ability to differentiate different states is not enough for the emergence of consciousness: the system must also be able to integrate information. This postulate is motivated by the fact that, under non-pathological conditions, the phenomenological experience does not occur in a fragmented way, i.e., we do not experience the colors of objects separated from their shapes.

Using concepts from mathematical information theory, Tononi proposes a measure of consciousness, Φ. In IIT the integrated information Φ is an information measure of the repertoires generated by the whole system, compared to the repertoires generated by the subsystems. Φ is defined in such a way that one of its consequences is the possibility of existence of different levels of consciousness. The set of elements within a system endowed with this property to generate a local maximum of conceptual information integrated is called complex.

For Tononi, systems that are able to generate consciousness are made of sub-mechanisms, and those sub-mechanisms can be combined in different ways to create mechanisms. It is the particular configuration of sub-mechanisms, at a given moment, that Tononi calls "context."

Though Tononi clearly sees the relevance of contexts to consciousness, in his model he makes the following assumption for the outcomes of mechanisms (see the supplementary methods of [20]):

$$p\left(ABC^t|ABC^{t-1}\right) = p\left(A^t|A^{t-1}\right) p\left(B^t|B^{t-1}\right) p\left(C^t|C^{t-1}\right). \tag{2}$$

Tononi's justification for (2) is that there are "no instantaneous interactions between mechanisms and causes precede their effects." It also seems to be an practical essential assumption to allow for computations in his model.

First, we should point out that the appeal to instantaneous interactions is misleading. To rule out instantaneous interactions, one would have to assure that the observations corresponding to, say, A^t or B^t, were separated by an spacelike interval (see [3] for a detailed argument in a different context). For example, if A^t and B^t were measured for an amount of time δt and were processes situated at a distance d, then any (non-instantaneous) interaction whose propagation speed is lower than the speed of light could account for violations of (2) if $\delta t \geq cd$. Since typical distances within the brain are at the order of 10^{-1} m, this means that for processes taking longer than 3×10^9 s, we can always explain them with non-instantaneous signaling. But most cognitive processes are believed to take much longer than 10^8 s. So, for biological processes, it is quite reasonable to assume that (2) may be violated due to physically plausible interactions between mechanisms (see [4] for a simple neural oscillator model exhibiting contextuality). Furthermore, we should point out that (2), as shown by Suppes and Zanotti [24], implies the existence of a jpd. Therefore, the assumption behind (2) is not, as Tononi claims, that of no instantaneous interactions: it is, instead, an assumption about *no* contextuality!

To show how contextuality may appear in IIT, let us focus on the mechanisms shown in Figure 1. This system behaves in a very simple way, and it is

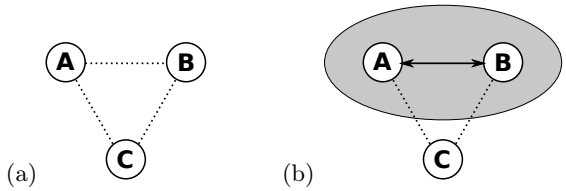

Fig. 1. (a) Contextual system of mechanisms, composed of six sub-mechanisms, A, B, and C. The sub-mechanisms are such that only pairs are simultaneously observable. (b) For example, when A is active, so is B, but not C. This is shown in the figure by the grayed area for the system.

constructed merely to show how contextuality can emerge here. We start with three mechanisms, A, B, and C, each taking values ± 1, which we represent by the ± 1-valued random variables **A**, **B**, and **C**. Let us assume that those mechanisms are stochastic, but, more importantly, that we can only observe the following simultaneously: (A, B), (A, C), (B, C), and (A, B, C). If contextuality is present, it is possible to have pairwise correlations for the situations where (A, B), (A, C), (B, C) are observed such that (1) is violated. In other words, any measurements except (A, B, C), give strong negative correlations between **A**, **B**, and **C**. This mechanism is explicitly contextual, since the marginals from (A, B, C) cannot match the correlations from the pairwise observations, and it is also implicitly contextual, since the pairwise correlations lead to no jpd. Now, let us imagine that, in this case, the mechanism is such that at time t only one of the pairs (A, B), (A, C), (B, C) is observed, and at time $t - 1$ the triple mechanism (A, B, C). It is clear, in this case, that equation (2) cannot hold, since there is no joint probability distribution.

One might argue that no mechanism is truly contextual, since we could, in principle, measure all quantities of interest simultaneously. This is not necessarily true for the following reasons. The first one is that in some cases it is not, *in principle,* possible to measure all quantities simultaneously. If, and we are not making this case here, there are underlying processes in the mechanism that are quantum, non-commuting observables cannot be simultaneously measurable, and contextuality may exist. This is what happens with entangled states in quantum mechanics, as in the famous Bell-EPR setup.

The second reason, which we consider more relevant, is simply empirical. It may be true that, for a certain system, it is possible *in principle* to measure all relevant variables simultaneously. However, for reasons of experimental limitations, it is close to impossible to measure them all. For example, imagine, that A, B, and C are neural oscillators that are measured with EEG. Because the activity strength of neural oscillators varies, and because EEGs are not spatially localized and are noisy, it is possible that under a certain stimulus, the only

observable neural oscillators are a subset of the relevant oscillators involved in the mechanism (say, A, and B). This does not necessarily mean that the other oscillator C has no existing outcomes, but simply that it cannot be measured under the experimental conditions. As a result, because of contextuality, when correlations are observed, they are enhanced by the selection of a subset of oscillators, and no jpd compatible with the observations will exist, even though all quantities are in principle well defined simultaneously. This, of course, would provide an empirical difficulty to guarantee that the system is not contextual.

Finally, the third reason is that, for some contextual systems, the experiment of measuring the pairs (A, B), (A, C), (B, C) has different marginal expectations (for correlations) than what you would observe for the triple (A, B, C). In this case, one could have direct influences from context, as the marginal expectations change. We emphasize that such direct influences do not imply any non-local interactions, since, as argued above, their time scales are large compared to the distance scales. To summarize, we provided here a toy example showing how context-dependent mechanisms violate (2).

4 Conclusions

IIT is a remarkable theory that opens up the possibility of empirically measuring consciousness. As such, it has the potential to have a significant impact in the way we think about consciousness. This explains how IIT has become one of the main theories in consciousness study.

One of the main ideas in IIT is that consciousness comes from processes that integrate information. Therefore, to measure consciousness, one needs to be able to measure the integration of information, and, more basically, information itself. The question is how to measure information for contextual systems, in light of the trilemma we mentioned at the outset. We know, for instance, that Shannon's entropy is not adequate for some contextual systems, and in particular, we know that for the special type of contextuality constrained by the formalism of quantum mechanics, the more appropriate measure of information is given by von Neumann's entropy. However, a more general way of measuring information in more general contextual systems, such as those necessary for the description of consciousness, is yet to be developed.

Until such measures of information in contextual systems are developed, the use of IIT to measure consciousness needs to be clarified, in particular as to how it is to be applied to contextual systems, which candidate systems to consciousness are believe to be.

Acknowledgments JAB and LPGA acknowledge support from the Patrick Suppes Gift Fund at Stanford University. Part of this work was developed at the Suppes Brain Lab in the Center for the Study of Language and Information, CSLI, and we thank CSLI and Professor John Perry for their hospitality.

References

1. Samson Abramsky and Lucien Hardy. Logical Bell inequalities. *Physical Review A*, 85(6):062114, June 2012.
2. Diederik Aerts, Jan Broekaert, Liane Gabora, and Tomas Veloz. The Guppy Effect as Interference. In Jerome R. Busemeyer, François Dubois, Ariane Lambert-Mogiliansky, and Massimo Melucci, editors, *Quantum Interaction*, number 7620 in Lecture Notes in Computer Science, pages 36–47. Springer Berlin Heidelberg, 2012.
3. J. A. de Barros and P. Suppes. Quantum mechanics, interference, and the brain. *Journal of Mathematical Psychology*, 53:306–313, 2009.
4. J. Acacio de Barros. Quantum-like model of behavioral response computation using neural oscillators. *Biosystems*, 110(3):171–182, December 2012.
5. J. Acacio de Barros, Janne Kujala, and Gary Oas. Negative Probabilities and Contextuality. *arXiv:1511.02823 [physics, physics:quant-ph]*, November 2015. arXiv: 1511.02823.
6. J. Acacio de Barros and Gary Oas. Some Examples of Contextuality in Physics: Implications to Quantum Cognition. In Ehtibar Dzhafarov, Ru Zhang, and Scott M. Jordan, editors, *Contextuality From Quantum Physics to Psychology*. World Scientific, 2015.
7. E. N. Dzhafarov, Ru Zhang, and Janne Kujala. Is there contextuality in behavioural and social systems? *Phil. Trans. R. Soc. A*, 374(2058):20150099, January 2016.
8. Ehtibar N. Dzhafarov and Janne V. Kujala. Quantum Entanglement and the Issue of Selective Influences in Psychology: An Overview. In Jerome R. Busemeyer, François Dubois, Ariane Lambert-Mogiliansky, and Massimo Melucci, editors, *Quantum Interaction*, number 7620 in Lecture Notes in Computer Science, pages 184–195. Springer Berlin Heidelberg, January 2012.
9. Ehtibar N. Dzhafarov and Janne V. Kujala. All-Possible-Couplings Approach to Measuring Probabilistic Context. *PLoS ONE*, 8(5):e61712, May 2013.
10. Ehtibar N. Dzhafarov and Janne V. Kujala. Contextuality in Generalized Klyachko-type, Bell-type, and Leggett-Garg-type Systems. *arXiv:1411.2244 [physics, physics:quant-ph]*, November 2014. arXiv: 1411.2244.
11. Ehtibar N. Dzhafarov and Janne V. Kujala. Context-Content Systems of Random Variables: The Contextuality-by-Default Theory. *arXiv:1511.03516 [quant-ph]*, November 2015. arXiv: 1511.03516.
12. Ehtibar N. Dzhafarov and Janne V. Kujala. Conversations on Contextuality. In Ehtibar Dzhafarov, S. Sordan, Ru Zhang, and Victor Cervantes, editors, *Contextuality from Quantum Physics to Psychology*, volume 6. World Scientific Press, New Jersey, 2015. arXiv: 1508.00862.
13. Ehtibar N. Dzhafarov, Janne V. Kujala, and Jan-Åke Larsson. Contextuality in Three Types of Quantum-Mechanical Systems. *Foundations of Physics*, 45(7):762–782, March 2015.
14. E.N. Dzhafarov and J.N. Kujala. A qualified Kolmogorovian account of probabilistic contextuality. *Lecture Notes in Computer Science*, 8369:201–212, 2014.
15. A. Fine. Hidden Variables, Joint Probability, and the Bell Inequalities. *Physical Review Letters*, 48(5):291–295, February 1982.
16. E. Haven and A. Khrennikov. *Quantum Social Science*. Cambridge Univ. Press, Cambridge, 2013.

17. Andrei Khrennikov. Ubiquitous Quantum Structure: from psychology to finance. 2010.
18. A.N. Kolmogorov. *Foundations of the theory of probability*. Chelsea Publishing Co., Oxford, England, 2nd edition, 1956.
19. Carlos Montemayor and Harry Haroutioun Haladjian. *Consciousness, Attention, and Conscious Attention*. MIT Press, April 2015.
20. Masafumi Oizumi, Larissa Albantakis, and Giulio Tononi. From the Phenomenology to the Mechanisms of Consciousness: Integrated Information Theory 3.0. *PLoS Comput Biol*, 10(5):e1003588, May 2014.
21. Daniel N. Osherson and Edward E. Smith. On the adequacy of prototype theory as a theory of concepts. *Cognition*, 9(1):35–58, January 1981.
22. Emmanuel M. Pothos and Jerome R. Busemeyer. Can quantum probability provide a new direction for cognitive modeling? *Behavioral and Brain Sciences*, 36(03):255–274, 2013.
23. Robert Stalnaker. Context and content: Essays on intentionality in speech and thought. 1999.
24. Patrick Suppes and Mario Zanotti. When are probabilistic explanations possible? *Synthese*, 48(2):191–199, 1981.
25. Giulio Tononi and Christof Koch. Consciousness: here, there and everywhere? *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 370(1668):20140167, May 2015.
26. Zheng Wang and Jerome R. Busemeyer. A Quantum Question Order Model Supported by Empirical Tests of an A Priori and Precise Prediction. *Topics in Cognitive Science*, 5(4):689–710, October 2013.
27. Zheng Wang, Tyler Solloway, Richard M. Shiffrin, and Jerome R. Busemeyer. Context effects produced by question orders reveal quantum nature of human judgments. *Proceedings of the National Academy of Sciences*, 111(26):9431–9436, July 2014.