Integrated Information Theory as Formal Framework for the Gradation of Social Structure

Benjamin Christensen

Department of Philosophy, Aarhus University, 8000 Aarhus C, Denmark

Filbc@cas.au.dk

Introduction

Social ontology concerns the investigation of such entities as rituals, institutions, conversations, riots, norms, artworks, corporations, language, families, and laws. Do they exist? If so, in what sense? Do they exist like stones or fish? If not, in what other sense do they exist? Such are customary examples of the questions being asked by social ontologists. Standardly, one of the most central disputes within the sub-discipline has been the individualism-holism debate: ontological individualists argue that social entities are fully ontologically exhaustible by the individuals composing them, while ontological holists maintain that social entities are ontologically irreducible, (Cf. Epstein 2018).

My aim in this paper is to argue that it is possible to construe the relevant social phenomena instead with gradualist terminology—and, indeed, that social ontology will be more fruitful and potentially have better contact with contemporary social science desiderata if terminologically recast in this way. I will therefore in the following not be concerned with presenting arguments for, or rebutting objections to, either ontological individualism or holism about social entities; instead, I will sketch a new approach to the formalization of our intuitions and claims about social entities.

Specifically, I propose that certain formal tools extractable from Integrated Information Theory (IIT) can provide us with a suitable formal language for the description and analysis of social entities. While IIT was originally developed for the specialized task of formalizing neural correlates to conscious states in the human brain, it has in recent years increasingly been reconsidered as a general-purpose framework for the formal description of system traits currently sparking interest within virtually all the sciences, such as robustness, structure, and emergence (Hoel et al. 2013; Holovatch et al. 2017; Marshall et al. 2018; Gomez et al. 2021; Grasso et al. 2021; Mediano et al. 2022). Remaining under the impression that IIT continues to be a theory with exclusively neuroscientific applications thus is to misunderstand the status-quo of the theory's development in the scientific community. To be sure, innovative developments within the specialized sciences—even if exploratory in nature—are of potential relevance to philosophers in general and philosophers of science in particular. Additionally, although I limit the discussion here to the potential usefulness of IIT as a formal framework for addressing questions in social ontology, it is worth noting that the method stipulated can plausibly be extended to other philosophical contexts pertaining to the relationship between parts and wholes.

Currently, almost no attempts have been made to employ the terminology of IIT within the social sciences, with a 2018 study by David Engel and Thomas W. Malone as one notable exception. Using an integrated information measure,[1] Engel and Malone calculated the integrated information of three types of social phenomena as modelled in datasets from empirical studies (cf. 2018): i) small groups of four persons collaborating to solve different tasks, ii) co-editing of Wikipedia articles, and iii) activity on a subset of the internet (ibid., 5-7). They found significant correlations with degrees of integrated information for all data-sets—between integrated information and the collective intelligence of groups from data-set (i), between integrated information and groups editing higher quality articles on Wikipedia from dataset (ii), and between increased integrated information and increased activity on a subset of the internet over time from dataset (iii) (ibid).

Finally, note that I will not here go into detail with the mathematics behind the integrated information measure. Its modus operandi can and have been illustrated using graphics consisting of interconnected nodes equipped with transition functions, and the calculations for our toy examples can be made with the Python "PyPhi" library developed for that purpose by W. G. P. Maynor and colleagues (2018). Really, the fact that the formal tools presented in this paper can now be utilized through programming is in itself reason for attention, since this potentially allows for analysis of objects conventionally far too complex for detailed formal philosophical discussion.

The paper is structured as follows. In the first section, the individualism-holism debate in social ontology is discussed. As opposed to the sum-zero approach whereby social entities are argued to be either wholly reducible or irreducible to individuals, a gradience approach as suggested by Harold Kincaid is highlighted. Although such an approach is promising, precise formal tools for the analysis of purportedly different gradations of social entity irreducibility are lacking. In sections II through V I first introduce IIT, and argue thereafter that the formal tools needed for the logically precise analysis and discussion of gradations of social entity irreducibility can be extracted from this theory. In section VI, I discuss one possible objection to the application of IIT proposed in this paper, and highlight

[1] To be precise, they used a less computationally demanding alternative developed by A. B. Barrett and A. K. Seth (2011).

in extension one live avenue for future research into the philosophical assumptions behind IIT.

I. Social Ontology and the Individualism-Holism Debate

According to ontological individualism, any social entity S would be exhaustibly replicated if all individuals involved in its composition were copied *simpliciter* (Epstein 2009, 187-89). Some ontological individualists construe this determination relation as a case of ontological reduction according to which social entities such as nation-states, international corporations and riots are nothing over and above the sum of the individuals composing them (Quinton 1976). Quinton, e.g., writes:

> I believe, then, that all statements about social objects are statements about individuals, their interests, attitudes, decisions and actions. But the predicates of these statements about individuals will essentially, if only implicitly, mention social objects in a way that is not practically or usefully eliminable, even if it is eliminable in principle (1976, 25).

Others have proposed less stringent determination relations such as supervenience (Searle 1995; Pettit 2003; Cf. Epstein 2018, 3.1). Pettit, e.g., writes:

> [I]f we replicate how things are with and between individuals in a collectivity—in particular, replicate their individual judgments and their individual dispositions to accept a certain procedure—then we will replicate all the collective judgments and intentions that group makes (2003, 184).

Another significant point of internal divergence among ontological individualists is the question of what counts as parts of social entities (Epstein 2018, 3.1). Some have argued that social entities are ontologically reducible to or supervenient upon certain psychological states of the individuals involved (Quinton 1976; Bratman 1999; Cf. Epstein 2018, 3.1.1). Others have opted beyond the psychological domain, suggesting that social entities in addition reduce to or supervene upon e.g. individual actions, relations, and/or adherence to norms (Kincaid 1986; Searle 1995; Cf. Epstein 2018, 3.1.1).

Ontological holism is the converse position relative to ontological individualism, according to which no social entity S would be exhaustibly replicated if all individuals I_1, I_2, I_3 … I_n involved in its composition were copied simpliciter. While it is rare to see contemporary philosophers of social science endorsing strong versions of ontological holism committed to an outright ontological priority of social wholes over their individual parts, it is, so Brian Epstein has argued, quite commonly held that there are principal obstacles "to the reduction of social phenomena to individualistic ones, even though the social is exhaustively determined by the individualistic" (2018, 3.3.1). Indeed, Epstein singles out many central voices within the field as opting in one way or other over the last 40 years for this middle-ground "non-reductive individualism" inspired by the analogous compromise of non-reductive physicalism within the philosophy of mind (ibid.), from Pettit (1981), Mellor (1982), Currie (1984), Kincaid (1986), and Tuomela (1989) in the 1980's, through Little (1991), Bhargava (1992), Hoover (1995), and Stalnaker (1996) in the 1990's, to Sawyer (2002) and List & Spiekermann (2013) since 2000 (Cf. Epstein 2018, 3.3.1).

At this point, the hard distinction between ontological individualism and holism is blurred (cf. Kincaid 2014). It might reasonably be asked what exactly individually adhering to, say, norms would entail. Norms, the ontological holist will object, inherently entail something over and above individuals. Recently, instead of doubling down on the distinction by broadening the scope of one category to include the terms of the other, these tendencies have been argued by some to signal the insufficiency of the dichotomy itself. Harold Kincaid exemplifies this approach when he argues that

> the live issues in the individualism-holism debate are not global ones to be decided on general conceptual grounds but local and contextual empirical debates about how far we can get by proceeding without institutional and social detail (2014, 140).

Instead, he continues, emphasis should be put on *how much* social structure a purported social entity has:

> There are numerous ongoing social science controversies that can reasonably be seen as instances of an individual-holism debate. Most of these can be phrased as debates about how holist or individualist can or must we be? This question is roughly about how much

> social structure we need to add to facts about individuals in order to successfully explain social phenomena? I take "social structure" to run the gamut from relatively thin social roles—a is the neighbor to b—to the full-fledged invoking of large scale social entities, e.g. nation states maximizing their interests in international relations with other nation states, with lots of mixes in between these two extremes. I think there are multiple cases where the question how holist or individualist we must be are core empirical issues in the social sciences (2014, 147).

In other words, while either a bank or a ritual may be examples of social entities, they are not necessarily on the same footing vis-à-vis their reducibility to the individuals respectively composing them. Furthermore, different instantiations of the same social entity type may vary greatly. Rituals, for one, span from e.g. local family traditions to large-scale procedures involving thousands of individuals at a time, repeated for hundreds of years, such as the Eleusinian Mysteries. Once considered, these factors of granularity and gradience seem unavoidable for just about any social entity conceivable. Kincaid provides several examples to substantiate this point, e.g. rational choice game theory:

> Many realistic games that individuals must play turn out to have multiple equilibria, given the constraints assumed on preferences, payoffs, and the like. How are such multiple possibilities reduced to one unique outcome in reality? For many games there seem to be focal points—socially salient choices—that allow people to coordinate. These focal points come from social norms and institutional arrangements that are not derived from the rational choice game theory models. So in this sense rational choice game theory must on occasion be even more "holist" (ibid., 147-48).

In line with the causal overriding criterion (Zahle & Collin 2014, 3), Kincaid proposes that the degree of social structure that we must take into account along with facts about individuals can be quantified as the magnitude of the social entity's causal effect over and above that of the individuals (2014, 150). This raises the issue that Kincaid proposes to consider as particularly germane over and above the individualism-holism dichotomy itself. Namely, given that we view social entities as possibly more and less social, how do we describe the weight of different social factors' causal effects? And how do we relate social structure to individuals? Novel approaches to such gradience and granularity issues, Kincaid argues, are

more properly the desiderata of the social sciences today than global arguments in favor of either individualism or holism. The most fruitful future course of action, Kincaid concludes, is to develop "integrated interlevel accounts" of individual and social entities, e.g. through the use of multilevel structural equation models (ibid., 150-51). That is, the direction projected is towards the use of *formal* tools for the analysis of the *degrees* of social structure associated with each purported social entity.

On the face of it, an integrated interlevel account of individual and social entities as stipulated by Kincaid would indeed avoid the limitations associated with globally holist or individualist approaches. However, in order to properly describe and analyze different gradations of social structure associated with different local examples of social entities, precise formal tools are needed that Kincaid does not offer in his paper. In the remaining sections, I provide first a general introduction to IIT, and argue then that the precise formal tools needed for an integrated interlevel account of individual and social entities can be extracted from this formal language.

II. What is Integrated Information Theory?

While Integrated Information Theory was originally devised for the specific purpose of formalizing, quantifying, and predicting different states of conscious states in human brains relative to physical, i.e., neural, correlates, researchers from fields ranging from mathematics (Tegmark 2016) through quantum physics (Zanardi et al. 2018), computer science (Mediano et al. 2022) and biology (Niizato et al. 2018), to social science (Engel & Malone 2018) have utilized its formal tools for problem-solving within their respective domains irrespective of its success or failure as a neuroscientific theory (Cerullo, 2015). One of the main drivers of the disciplinarily wide-ranging interest in the formal tools of IIT—and the one that is of specific interest to us for our present purposes—is that they characterize any system that they are applied to relative to its degree of functional integration. To be sure, while the type of system that were originally intended to be characterized in this manner was the human brain, there is no principal obstacle to conceiving systems in general after a similar fashion. As a first approximation to what is meant by characterizing systems *in general* relative to their degrees of functional integration we may therefore look to the example of Tononi's original application of the theory.

In brief, IIT was initially intended as a conceptual framework within which to account for empirical findings concerning the relationship between the human brain and conscious states as a function of its activities that do not neatly fit into a linear picture of that type of system. For instance, Tononi originally used the formal tools of IIT to argue that rather than the sheer number of neurons in the system, what significantly correlates to the function "conscious" is the degree to which the neurons process information in an interconnected way (2004, 221). Such an approach, so he contended, can account for the empirically well-founded assumption that the cerebral cortex, which, as he points out, contains many "highly specialized elements" that fulfil "unique functional role[s]" and "many path-ways for interactions among the elements", is a much more relevant neurological correlate to conscious states in the overall system than the functionally simpler cerebellum, which however contains the greater number of neurons (ibid.). "Integrated information", thus, is supposed to capture

> how much information is generated by a single entity, as opposed to a collection of independent parts. The idea here is to consider the parts of the system independently, ask how much information they generate by themselves, and compare it with the information generated by the system as a whole (2004, 221).

Where "information" as a first approximation can be thought of like Shannon information, i.e., signifying how much of what we know about a system or sub-system in their current state constrains what we can know about their previous and future states (Marshall et al. 2018, 6-7).

However, as opposed to Shannon information, the term "information" is used in IIT in a causal sense, i.e., as the features of one physical system with which another physical system interacts. In this sense of the term, a thermometer has "information" about the room because its current state α is causally constrained by the room's temperature. *Informational* constraints, in short, are in IIT taken to be produced in physical systems by *causal* constraints (Oizumi et al. 2014, 6). What a system can do and what can be done to it, in turn, are taken in IIT to be all that there is to it, ontologically speaking. The core ontological assumption in IIT thus is the ontological principle of difference-making known from Plato's *Sophist* as the "Eleatic principle" (247d-e), namely that to exist is to make a

difference. Thus, IIT predicts that if a physical system such as an intact brain constrains at t_0 to a higher degree what we can know about its states at t_{+1} than the sum of its parts does, then this is because the intact brain makes a causal difference to itself over and above the parts, which is to say: it is an ontologically irreducible entity (Tononi 2012, 297).

Consider now, as opposed to a human brain, a system that intuitively generates highly similar or even identical information as compared to its parts. Tononi provides the example of photodiodes in a digital camera (ibid., 294). No lesion to one part of such a system will affect its other parts—that is, no removal of a single pixel will affect other pixels (ibid.). In other words, the sum information of the individual pixels composing a picture taken by the camera will be identical to the picture as a whole (ibid.). Luis H. Favela has afforded the useful terminology of "component" and "interaction" dominant systems to distinguish between these extremes (2019, 31). Our photography is an example of a "component dominant" system. The structure of the system is highly additive; it is simply the sum of its parts. Vice versa, the parts are not individually structured by the system as a whole (ibid.). This system, then, functions completely linearly (ibid., 35). Conversely, in an interaction dominant system such as the human brain the activity and organization of its parts will "supersede those that the parts would have separately from each other" (ibid., 33). The global activity of such a system will be sustained by the continuous interaction of its parts (ibid., 33-34). This results in a situation where there "is a global-to-local as well as local-to-global cause-effect relationship that maintains a particular order" (ibid.). Interaction dominant systems thus function nonlinearly with feedback loops, and tend to stabilize into robust patterns of behavior *resultant* upon these loops (ibid.).

Thus, generalizing IIT for our specific purposes, we will operate with the premise that if a social entity S makes no difference over and above the differences made by the individuals involved in it, then it is nothing over and above them, and we can be ontological individualists as far as it is concerned. More importantly, the integrated information measure may not only be utilized to decide disputes about the ontological irreducibility of specific social entities on a sum-zero basis, but also for distinguishing between differing gradations of social structure, as appealed for by Kincaid.

III. Calculating Integrated Information[2]

It should now be clear that far from being restricted to the description of neural correlates, integrated information can be conceived much more generally to potentially capture the degree to which *any* entity makes a difference over and above its parts. I pass on now to detailing more carefully the *modus operandi* of IIT's formal tools.

In the conceptual framework of IIT, a physical system is formally defined as

> a set of elements, for example neurons in the brain or logic gates in a computer, such that each element has at least two states, inputs that can influence these states, and outputs that in turn are influenced by these states (Marshall et al. 2018, 6).

In order to capture this, IIT represents physical systems as networks of interconnected nodes that enter either of two mutually exclusive states at any timestep t+1 based on the states of their inputs at time t (Mayner et al. 2018, 2; Mayner et al. 2018, S1, 4), as illustrated in figure 1 below (Cf. Mayner et al. 2018).

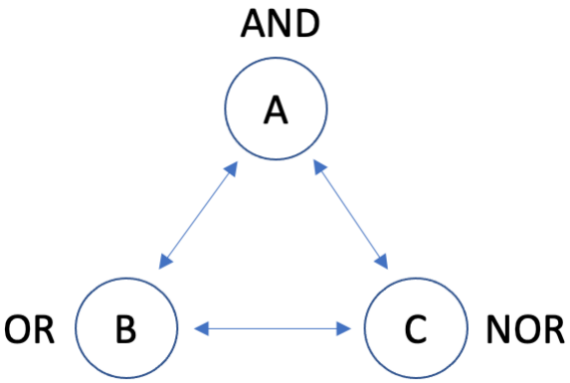


Figure 1.

A system of interconnected nodes, here illustrated with logic gates for the transition functions. A will only have the value "on" if B and C were "on" at the previous timestep t-1, B only if A, C, or both were "on" at t-1, and C only if neither A nor B were "on" at t-1. Note that the transition functions do not need to be those of logic gates; these are just convenient for the sake of illustration.

[2] I follow here mainly the line of expositions provided by Oizumi and colleagues in their 2014, and Mayner in the PyPhi material at https://pyphi.readthedocs.io/en/latest/examples/2014paper.html, which is again based on Oizumi and colleagues' 2014. Since our purposes here are different, many details are left out for the sake of simplicity. Readers interested in the mathematical details are referred to Oizumi and colleagues' 2014 paper; readers interested in Mayner's comprehensive exposition of the use of PyPhi for calculating integrated information are referred to the documentation at: https://pyphi.readthedocs.io/en/latest/.

With this definition of physical systems IIT adopts the position of "operational physicalism" as implied by the difference-making principle (Albantakis et al. 2022, 6). According to operational physicalism the only way for us to know whether something exists is to determine through systematic observation and manipulation whether or not it makes a difference and a difference can be made to it (ibid.). In order to observe and manipulate physical systems supported by precise mathematical notation, IIT uses transition probability matrices to describe their causal constraints, since transition probability matrices provide information about transitions between states within a system (ibid.). Recall that information is in IIT construed as causal constraints: to restate, similarly to how the "information" in a thermometer is equal to the causal constraints of the surrounding space upon it, information about state transitions in physical systems and their functional parts is in IIT taken to be equal to the mutual causal constraints of these.

Thus, physical systems are in IIT represented as networks of interconnected nodes, where each node can be in (at least) two possible (mutually exclusive) states, and the states of the nodes are determined by transition functions from past to future states relative to input (Mayner et al. 2018, 2; 5). A physical system, then, is characterized by (i) the states that it is possible for its nodes to currently be in, (ii) the logically possible previous states given any current state S, and (iii) the logically possible subsequent states given any current state S. This can be illustrated as in the example system E in figure 2 below (Cf. Mayner et al. 2018, S1):

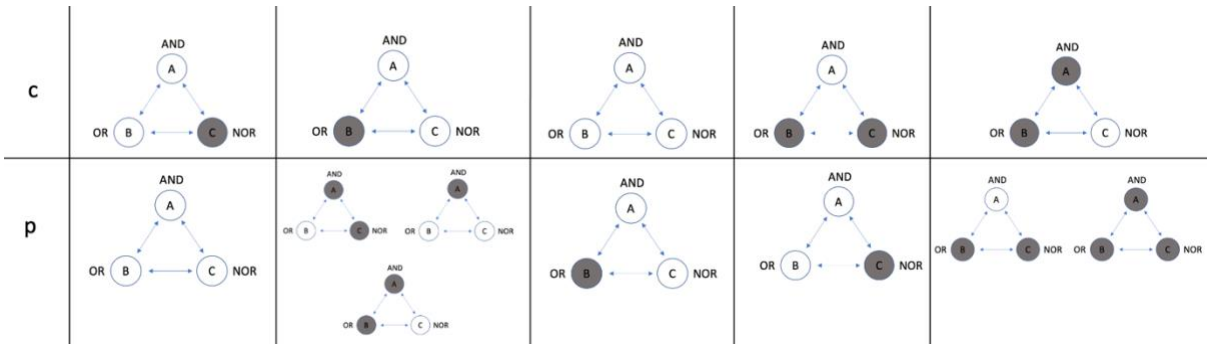


Figure 2.

Example system "E". In IIT, physical systems are represented as interconnected nodes with transition functions from past to future states. The rubric "c" stands for "current state" and the rubric "p" stands for "past state" (Oizumi et al. 2014, 6).

In the rubric "c", this illustration shows the set of states that are possible for the example nodes from figure 1 to enter into given the specified transition functions. If e.g. only the NOR-gate C is "on" at c, then we know that at p, neither A nor B nor C were "on", since that is the only possible state to have preceded that state. On the other hand, if the AND-gate A and the OR-gate B is both "on" at c, then we are limited to knowing that at p, either B and C were both "on" with A "off", or A, B, and C were all "on". The Boolean states are transcribed using 1's and 0's, so that e.g. the last reads rubric reads: $ABC^c = 110$ constrains the past possible states of the system to $ABC^p = 011$ or $ABC^p = 111$ (Cf. Oizumi et al. 2014, 6). Note that we must equally consider the possible future states "f" of a system given any current state c. For instance, a current state $ABC^c = 000$ constrains the past state of the system E to $ABC^p = 010$, but the future state of the system to $ABC^f = 001$.

The integrated information of a system is then calculated over two main steps, the first of which calculates integrated information for the individual functional units of a system S, the second for the overall system (Oizumi et al. 2014, 4-10; 10-14). The integrated information of individual functional units is signified by lower case "ϕ" while upper case "Φ" signifies system-wide integrated information (ibid.).

Turning to the first step, we may begin by noting that the functional units for which integrated information is here calculated are in IIT termed "mechanisms"[3] (Oizumi et al. 2014, 4; cf. Albantakis et al. 2022, 10). More specifically, the term "mechanism" signifies a subset of a system S that causally affects and is causally affected by another subset of S, termed a "purview" (ibid., 7). The subsets under consideration correspond to the powerset of S, such that if the simplest functional units of S are A, B, and C, then A, B, C, AB, AC, BC, and ABC are all candidates for mechanisms and purviews (ibid.). For instance, in the example system E illustrated above "BC" over the purview "ABC" is a mechanism given the state $ABC^c = 011$.

Next, we note that mechanisms are quantified as the information that they specify in the transition probability matrix which is used to describe the system (Oizumi et al. 2014, 6). They specify information at t about a given purview at t-1 by how they are affected by it, and at t about a given purview at t+1 by how they affect it (ibid.).[4] The information specified by a mechanism is equal to the distance D[5] between the probability distribution of the

[3] Although here kept to a minimum, there will be some terms with specialized uses in IIT in the following, such as "mechanism", "purview", and "minimum information partition". For a comprehensive overview of the key specialized terms used in IIT, see the 2014 of Oizumi and colleagues, 4.

[4] These are called the "cause repertoire" and "effect repertoire" of the mechanism, respectively. For technical reasons only the minimum of the two probability distributions is used for the next steps of calculation, but this need not occupy us here, as our agenda is not the technicalities of the information measure. For more detail on this, see e.g. Oizumi et al. 2014, 6.

[5] IIT uses the "Earth Mover's Distance" metric for this (ibid., 4). This metric measures the minimum cost of transforming states into other states (ibid., S2, 5). It works by multiplying the number of states transformed by

potential past and future states of its purview *as constrained by the mechanism at t* and the probability distribution of the potential past and future states of its purview *without any mechanism constraints* (ibid., 4-6). For instance, given the state $ABC^c = 011$, BC^c alone constrains the probability distribution to one option with probability 1, since the only possible state to have preceded $BC^c = 11$—the state of A^c notwithstanding—is $ABC^p = 001$. As an example of a mechanism that constrains to a lesser degree the probability distribution of the powersets of ABC^p and ABC^f, consider how mechanism AB given the state $ABC^c = 110$ constrains purview ABC^f. In this case, the mechanism AB constrains the purview ABC^f to either the state $ABC^f = 010$ or $ABC^f = 110$ with probability 0.5 (cf. ibid., 6). As an example of a value returned for information specified by a mechanism, calculation in PyPhi shows that for the state $ABC^c = 011$, the distance between the unconstrained probability distribution and the probability distribution of ABC^p as constrained by the mechanism BC has the value 1.5 (see appendix 1). This value is referred to as the "cause information" of the mechanism BC over the purview ABC^p, since it reflects the degree to which BC at t is causally constrained by ABC at t-1 (Oizumi et al. 2014, 6).

The integrated information of an individual mechanism is taken to be equal to how much information it generates over and above its parts (ibid., 4; 6-9). In order to measure this, the information specified by the intact mechanism is compared to the information specified by a partitioned version of it (ibid.). For instance, our example mechanism BC over purview ABC^p (BC/ABC) can be partitioned into mechanism B over purview C^p (B/C^p) and mechanism C over purview AB^p (C/AB^p) (ibid., 8). The product of the information specified by B/C^p x C/AB^p is then compared to the value of the information specified by BC/ABC^p to determine whether or not the mechanism BC over purview ABC^p generates information not reducible to information generated by its parts B over purview C^p and C over purview AB^p (ibid.; cf. Tononi 2012, 299). The integrated information ϕ of a mechanism is equal to the distance D between these values, which is calculated using the same distance metric as when calculating the information specified by mechanisms in the first place (Oizumi et al. 2014, 8).

---

the "Hamming distance" that they are moved. The Hamming distance between binary states corresponds to "the number of places by which two strings differ" such that "the Hamming distance between the states ABC = 000 and ABC = 111 is 3; the distance between ABC = 010 and ABC = 100 is 2" (ibid.).

Of course, if this comparison could freely be made between the intact mechanism in question and *any* partition of it, ϕ would be an arbitrary measure. To avoid this, IIT compares the information generated by a mechanism only to the partition of the mechanism that results in its *closest auxiliary in terms of information generated* (ibid., 4; 8). This partition is termed the "minimum information partition", since it is the version among all possible partitioned versions of the mechanism that suffers the *smallest decrease* in information generated as compared to the intact mechanism (ibid.). The integrated information ϕ of a mechanism, in short, is equal to the distance in terms of information generated between the intact mechanism and its minimum information partition (ibid.). Notice that if a mechanism does *not* generate integrated information, then that means that it does not make a difference over and above the differences made by its parts in the system, and therefore on pain of the difference-meaning principle *is not ontologically irreducible* (ibid., 8-10).

To recapitulate: In this step, subsets of a system S that are both causally affected by and causally affect other subsets of the powerset of S are identified. These subsets are termed "mechanisms". Then, those mechanisms are located that make differences not reducible to the differences made by their parts.

In the second step, the overall integrated information Φ for the whole system S is found largely by re-iterating the procedure for ϕ. That is, this step identifies differences made by S not reducible to differences made by *its* functional parts, namely the causally irreducible mechanisms located in the first step. This is because information about whole systems is according to IIT integrated across mechanisms; the more co-dependent the interactions between these mechanisms are, the more integrated the system will be (ibid., 10-14).

During this step the notion of a "minimum information partition" is once again used to identify the partitioned version of the whole system S that has the smallest decrease in information generated as compared to the intact system (ibid.). This time the partition is executed by performing a unidirectional cut to the interconnected nodes describing the system (ibid., 11-12). For instance, in the state ABC = 011, E has four causally irreducible mechanisms, namely A, B, C, and BC (see appendix 2). The minimum information partition of E in this state is a unidirectional cut that removes all connections from AC to B (see appendix 3), which leaves the system with only two causally irreducible mechanisms,

namely A and C (see appendix 4). The integrated information Φ is equal to the distance D between the information generated by the intact system and the information generated by the partitioned version of the system that results in the smallest reduction in mechanisms as compared to the intact one. E therefore is a causally irreducible and as such, according to the difference-making principle, *ontologically* irreducible system.

In short, we prima facie indeed have here a formal language that specifies not only *whether* the systems that it describes are ontologically irreducible, but also *the degree* to which they are so.

IV. Formalizing and Quantifying Social Structure as Integrated Information

In order to illustrate how the formal tools of IIT work, we will now discuss the steps introduced above using a deterministic[6] toy model of a simplified group cooperation scenario.

The toy model, call it "T", consists of three individuals, A, B, and C, where C must translate messages between A and B. We can describe this scenario as a system such that A can only activate B with the help of C, B can likewise only activate A with the help of C, and C only activates with either of A or B at a time, enforcing turn-taking, as illustrated in figure 3 below:

[6] Note that IIT is not limited to deterministic models. While I use here a deterministic toy model for the sake of simplicity, IIT is, due to its use of Markovian models for the representation of physical systems, perfectly geared towards the representation and analysis of nondeterministic systems (cf. Mayner et al. 2018, S1).

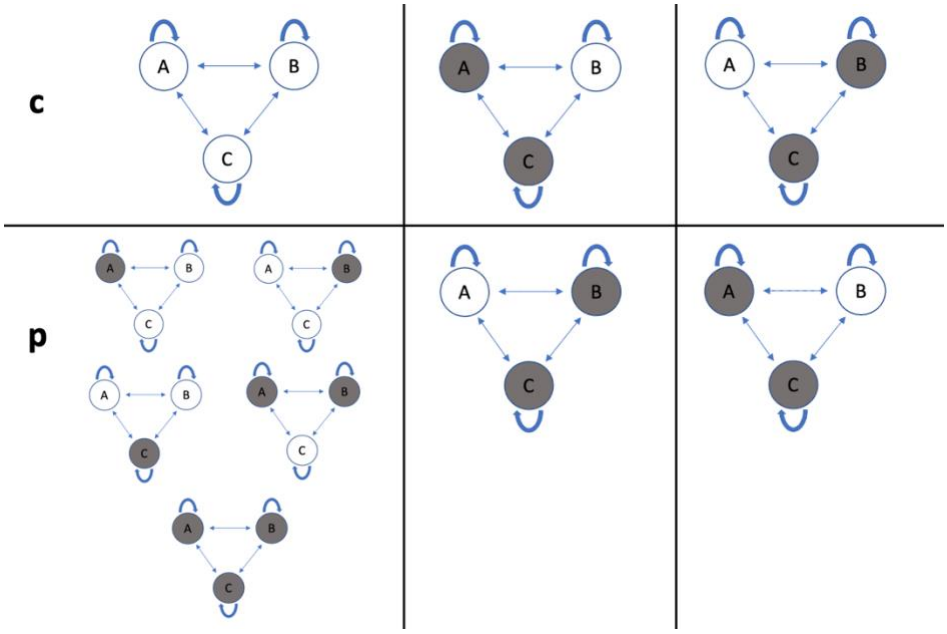


Figure 3.

As for the example system E in figure 2, the possible causal constraints of the different mechanisms involved in T with respect to the past of T are here represented by showing states at the timesteps "c" and "p". Note that T is not represented using logic gates, but highly specified transition functions: only if B is "on" together with C at p will A go "on" at c together with C, and vice versa. Note also that self-referential arrows have been introduced. This is done to simulate turn-taking; in order for B to keep up turn-taking and only go "on" when A and C were "on" at the previous timestep, it needs to know the difference between (i) B and C being "on" and (ii) A, B and C being "on"—and vice versa for A and C.

We turn now to repeating the two-step procedure detailed in the previous section, only this time specifically for our toy model T. As we saw above, the probability distributions of mechanisms A, B, C, AB, AC, BC and ABC over purviews A, B, C, AB, AC, BC and ABC must first be calculated. As another example, consider the situation for our toy model T in which B is currently cooperating with C, $ABC^c = 011$. $ABC^c = 011$ completely constrains ABC^p, since the only possible past state of ABC to have caused $ABC^c = 011$ is $ABC^p = 101$. This means that if $ABC^c = 011$ then $ABC^p = 101$ with probability 1. In plain, we can know that if C is currently cooperating with B, then A must just have cooperated with C.

As with the BC mechanism in system E used for exemplification in the previous section, we now calculate with PyPhi for T the distance between, on one hand, the past and future purviews ABC^p and ABC^f as constrained by mechanism BC and, on the other hand, the past and future purviews ABC^p and ABC^f considered *without* any mechanism constraining

them for the information specified by BC. This returns the values 1.5 for the information specified by BC over the past purview ABC^{p} and 0.75 for the information specified by BC over the future purview ABC^{f} respectively. As mentioned already in the footnotes above, only the minimum value between the mechanism over its past and future purviews is selected for the next steps of calculating integrated information.[7] The information specified by mechanism BC over purviews ABC^{pf} given the state ABC^{c} = 011 for system T, then has the value 0.75 (see appendix 5):

Information specified given state ABC^{c} = 011

BC over purview ABC^{pf}
0.75

Note that *all* mechanisms that specify information in T given the state ABC^{c} = 011 must be identified, as illustrated in figure 4 below (see appendix 6):

Information specified by all mechanisms over all purviews in T given state ABC^{c} = 011

A/A^{pf}	**A/B^{pf}**	**A/C^{pf}**	**A/AB^{pf}**	**A/AC^{pf}**	**A/BC^{pf}**	**A/ABC^{pf}**
0.071429	0.071429	0.0	0.142856	0.125	0.125	0.214284
B/A^{pf}	**B/B^{pf}**	**B/C^{pf}**	**B/AB^{pf}**	**B/AC^{pf}**	**B/BC^{pf}**	**B/ABC^{pf}**
0.125	0.125	0.0	0.25	0.125	0.125	0.25
C/A^{pf}	**C/B^{pf}**	**C/C^{pf}**	**C/AB^{pf}**	**C/AC^{pf}**	**C/BC^{pf}**	**C/ABC^{pf}**
0.0	0.0	0.25	0.25	0.375	0.375	0.5
AB/A^{pf}	**AB/B^{pf}**	**AB/C^{pf}**	**AB/AB^{pf}**	**AB/AC^{pf}**	**AB/BC^{pf}**	**AB/ABC^{pf}**
0.375	0.125	0.25	0.5	0.625	0.375	0.75
AC/A^{pf}	**AC/B^{pf}**	**AC/C^{pf}**	**AC/AB^{pf}**	**AC/AC^{pf}**	**AC/BC^{pf}**	**AC/ABC^{pf}**
0.071429	0.071429	0.166667	0.5	0.625	0.375	0.75
BC/A^{pf}	**BC/B^{pf}**	**BC/C^{pf}**	**BC/AB^{pf}**	**BC/AC^{pf}**	**BC/BC^{pf}**	**BC/ABC^{pf}**
0.375	0.125	0.25	0.5	0.625	0.375	0.75
ABC/A^{pf}	**ABC/B^{pf}**	**ABC/C^{pf}**	**ABC/AB^{pf}**	**ABC/AC^{pf}**	**ABC/BC^{pf}**	**ABC/ABC^{pf}**
0.5	0.125	0.5	1.0	1.0	0.875	1.5

Figure 4.

[7] Whenever I use both the "p" and "f" superscripts, this signifies the minimum value between the two that is used for further calculation.

A list of all mechanisms over all purviews in T given state ABC^c = 011. Note once again that this list includes for every purview only the minimum value between its past and future iteration, signified by the use of both the "p" and "f" superscript.

Then, in order to check if they specify information over and above their parts, all mechanisms are compared to their minimum information partition—just like the mechanism BC/ABC^p of example system E was compared to its minimum partition B/C^p x C/AB^p in the previous section. Once again, only the mechanisms that specify information over and above their parts have integrated information ϕ and are therefore causally irreducible functional units of T. In our case, four out of seven possible mechanisms generate ϕ in T given the state ABC^c = 011 (see appendix 7):

Mechanisms of T with integrated information (ϕ) in state ABC^c = 011

A	B	C	AC
0.071429	0.125	0.25	0.25

Note that if a mechanism does generate ϕ, it is counted at the purview where it generates *maximal* ϕ, i.e., where the distance between the information specified by the intact mechanism and the information specified by its minimum information partition is greatest, as illustrated in figure 5 below:

Purviews over which causally irreducible mechanisms are maximally irreducible

A/A^{pf}	A/B^{pf}	A/C^{pf}	A/AB^{pf}	A/AC^{pf}	A/BC^{pf}	A/ABC^{pf}
0.071429	0.071429	0.0	0.142856	0.125	0.125	0.214284
B/A^{pf}	B/B^{pf}	B/C^{pf}	B/AB^{pf}	B/AC^{pf}	B/BC^{pf}	B/ABC^{pf}
0.125	0.125	0.0	0.25	0.125	0.125	0.25
C/A^{pf}	C/B^{pf}	C/C^{pf}	C/AB^{pf}	C/AC^{pf}	C/BC^{pf}	C/ABC^{pf}
0.0	0.0	0.25	0.25	0.375	0.375	0.5
AB/A^{pf}	AB/B^{pf}	AB/C^{pf}	AB/AB^{pf}	AB/AC^{pf}	AB/BC^{pf}	AB/ABC^{pf}
0.375	0.125	0.25	0.5	0.625	0.375	0.75
AC/A^{pf}	AC/B^{pf}	AC/C^{pf}	AC/AB^{pf}	AC/AC^{pf}	AC/BC^{pf}	AC/ABC^{pf}
0.071429	0.071429	0.166667	0.5	0.625	0.375	0.75
BC/A^{pf}	BC/B^{pf}	BC/C^{pf}	BC/AB^{pf}	BC/AC^{pf}	BC/BC^{pf}	BC/ABC^{pf}
0.375	0.125	0.25	0.5	0.625	0.375	0.75
ABC/A^{pf}	ABC/B^{pf}	ABC/C^{pf}	ABC/AB^{pf}	ABC/AC^{pf}	ABC/BC^{pf}	ABC/ABC^{pf}
0.5	0.125	0.5	1.0	1.0	0.875	1.5

Figure 5.

Only the purviews over which causally irreducible mechanisms are *maximally* integrated are counted. For T in state ABC^c = 011, these are here highlighted in red.

In our case, ϕ is maximally generated by A over purview ABC^{pf}, by B over purview AB^{pf}, by C over purview C^{pf}, and by AC over purview AC^{pf}. Note that BC in fact is *not* counted among the mechanisms that generate information over and above their parts, i.e., in state ABC^c = 011 for our system T, BC is causally reducible.

Since a mechanism M exists if and only if it plays an irreducible causal role within a system S, the figures returned by the above calculations demonstrate (a) which, if any, possible subsets of our toy model actually have causal effects and are causally affected and are thereby to be counted as ontologically irreducible functional units of T, and (b) quantifies the *degree* to which they are causally irreducible. Note what this means in the more familiar terminology of social ontology broadly construed. If the composite *subsystem* entity of individuals B and C integrates information in a system-state, then, so goes the argument, it is a social entity over and above the individuals involved in that state, and has a certain degree of social structure. Of course, it may be objected that BC ought to be causally irreducible in all states of the toy model, but this would fail to connect to the overall point being made here. The contention is not that the particular toy model used here for the

representation of a cooperation scenario is necessarily the best representation of its explanandum, but simply that the formal tools of IIT presented provide an apt logical language within which to *have* such discussions. For instance, on the basis of the formalizations presented of T so far, it might be countered that the model in fact does capture the sort of structure intended, since the previous cooperation between A and C at $ABC^c = 011$ performs as an irreducible functional unit AC for the interaction currently going on between individuals B and C, which will at the next state $ABC^c = 101$ then perform as an irreducible functional unit BC for the interaction between individuals A and C.

Finally, we can with these formal tools consider whether T itself—under the premises of its representation and the difference-making principle—is ontologically irreducible and—if indeed it is—how much so. Φ, as we saw in the previous section, signifies system-wide causal irreducibility (ibid.). Similar to ϕ, Φ is found by comparing a maximally irreducible entity to its minimum information partition. This time, however, instead of individual maximally irreducible mechanisms, the entity at issue is the maximally irreducible *composite of mechanisms* within the overall system in question, which is termed the "major complex" of the system (Oizumi et al., 5). The greater the degree to which the minimum information partition of a system's major complex diminishes the system-wide causal profile, the more integrated is the overall system in question (ibid.). In the case of T, calculation in PyPhi returns a Φ value of 0.410218 and reveals that ABC composes a major complex, meaning the whole system T makes differences not only over and above A, B, and C individually, but also over and above AB, AC, and BC (see appendix 8):

System-wide irreducibility of T in state $ABC^c = 011$

T Φ	T major complex
0.410218	ABC

Accordingly, insofar as we accept the premise that the sort of interactions modelled by T can be construed as "social", the entity represented does according to the formal framework presented have irreducibly social structure, since it makes a difference over and above the differences made by the individuals composing it.

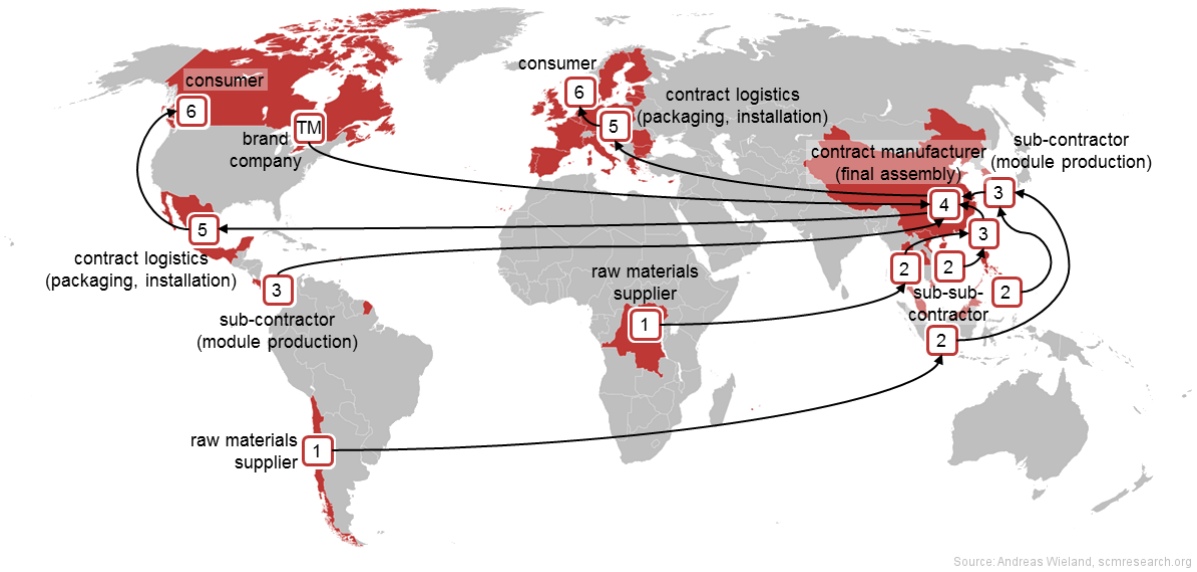
V. Comparing Social Structure

We have seen that the formal tools of IIT can principally be utilized for analytical work in the philosophy of social science in the following ways:

1) As a logical language in order to formalize our claims and intuitions about social entities.
2) As a tool for analyzing different elements of social entities, e.g., as seen above, (a) how many and what difference-making parts (i.e. mechanisms) the social entity as represented has, and (b) whether the social entity as represented can—premised upon the difference-making principle—meaningfully be said to be a distinct entity over and above its parts.
3) As a tool for quantifying different elements of social entities, e.g., as seen above, (c) *how much* of a difference, if any, the parts make in relation to each other, and (d) *how much* of a difference, if any, the social entity makes over and above its parts.

The last thing for us to consider is now how the framework can be used for comparison between different purported social entities.

For the comparison example, we will formalize and quantify within the logical language of IIT a simple existing social science model. The model is the following diagrammatic representation by Andreas Wieland (2018) of the computer industry supply chain:



As with our previous example, it is prima facie a plausible suggestion that the phenomenon modelled, i.e., a commercial system, is a social entity.

This, I take it, is exactly the sort of issue cautioned about by Kincaid. For such intuitive assessments can indeed easily slip into principled discussions about the ontological irreducibility of social entities in general rather than of the concrete case in question. In order to analyze the model within our formal language, then, we first formalize it according to the framework's principles. This could e.g. be done as in figure 6 below.

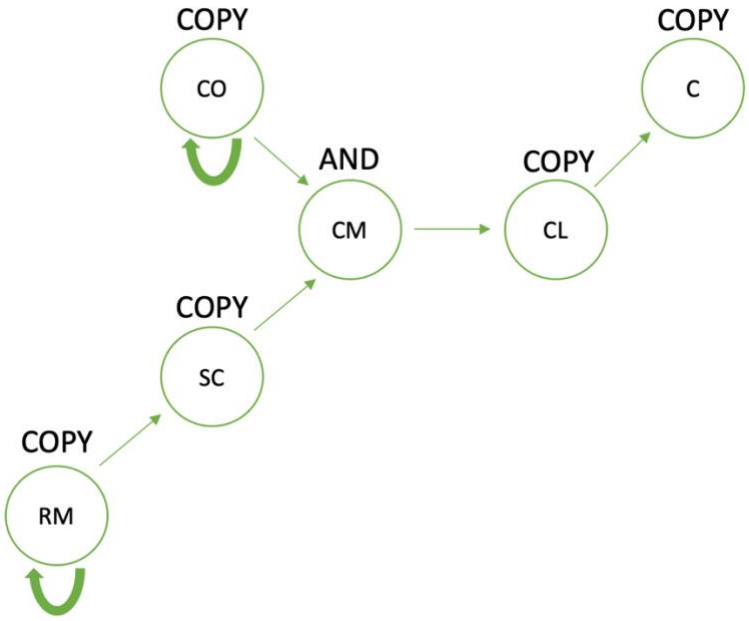


Figure 6.

The causal logic of Wieland's model of the computer industry supply chain as captured within the logical language of IIT. The contract manufacturer (CM) goes "on" at t only if both the brand company (CO) and sub-contractors (SC) were both on at t-1, the sub-contractors (SC) goes "on" at t only if the raw materials suppliers were "on" at t-1, contract logistics (CL) goes "on" at t only if the contract manufacturer (CM) was "on" at t-1, and finally consumers (C) go "on" at t only if contract logistics (CL) was "on" at t-1. Since neither of the brand company or the raw materials suppliers have any inputs in Wieland's diagram, they have simply here been set up with a recursive COPY function, so that either will go "on" at t only if it was itself "on" at t-1.

We can now check our second toy model social entity, call it "SU" for "supply-chain", for system-wide integrated information. If the model has integrated information, then we take this to indicate, as above, that it does indeed represent an entity with irreducible social structure at this level of organization. In this case, calculation with PyPhi (see appendix 8)

tells us that SU has no integrated information in e.g. state ABCDEF^{c} (i.e. CMRMCLCCOSC^{c}) = 111111:

System-wide irreducibility of SU in state ABCDEF^{c} (i. e. CMRMCLCCOSC^{c}) = 111111

SU Φ
0.0

Since SU does not have integrated information in any other state either,[8] we here have a case of a social-science model which could intuitively be argued to represent an ontologically irreducible social entity just as well as our toy model T above, but which, on the premises of the difference-making principle, is completely reducible to its parts, and therefore should not be counted as an ontologically irreducible social entity. Note that this says nothing about the potential social structures of the subsystems composing SU; each of these might very well on the same criteria be found to be ontologically irreducible social entities, even though they do not irreducibly integrate social information at the level of SU.

VI. Do individuals Exclude Social Entities According to Integrated Information Theory?

It seems indeed feasible, then, that a formal ontology based on the tools of IIT can provide an integrated levels account of individual and social entities able to book-keep causal interactions and distinguish between different grades of social structure among different social entities. Before concluding, one possible objection to this proposal merits consideration.

It could be objected to the above use of IIT's formal tools that social entities may, on the original version of IIT, be altogether excluded from the ontological record. In brief, this would be because the standard version of IIT as far as ontology goes operates with a winner-takes-all principle, called "exclusion", whereby for any system S, only the unit(s) with the highest integrated information are ontologically irreducible, meaning any other units, no matter their distance in integrated information to those with maximal integrated information, lose all ontological significance (Tononi 2012, 325). The objection, in short,

[8] I do not include an appendix demonstrating calculations in PyPhi for all possible states of SU because it would take up considerable space.

would be that if a social entity S contains any individuals I_1, I_2, I_3 … I_n, each with more integrated information than S, then S must be altogether ontologically excluded, with only the individuals remaining ontologically irreducible. This approach was imperative for the original formulation of the theory, as proponents needed to avoid the ad absurdum inviting scenario in which groups of conscious individuals—they name a conversation between two individuals and the United States of America as examples—are predicted to be themselves conscious (Tononi & Koch 2015, 13).

We can meet this objection, however, firstly by rejecting that integrated information is sufficient for consciousness. With this, the prima facie absurdity of conversations and nations genuinely exhibiting integrated information disappears, since the theory then no longer predicts that they are conscious. Instead, we may, as throughout this paper, hold simply that integrated information predicts ontological irreducibility.

To that it may be countered that IIT seems then to predict that only those sub- and supersystems in relation to a given system that have maximal integrated information are ontologically irreducible. This would once again problematize the notion that social entities could ever within the framework of IIT be considered ontologically irreducible, since integrated information for individual human beings—in particular because of the human brain—presumably outweighs any actual social entity.

But we may reply, then, that it is not necessary to assume that for any social entity S, any individuals I_1, I_2, I_3 … I_n involved are *completely* involved. Imagine the following scenario: A troop of human dancers perform Swan Lake. To be sure, if we could calculate the exact integrated information of individual human beings, it is quite conceivable that the integrated information of any individual would surpass the integrated information of any dancing performance. However, the objection presupposes that *every* feature that goes into the individual dancer's integrated information *is involved in the dancing*. At face value, we do not need to make this assumption; one of the dancers may have read, say, the collected works of Plato, but this does not necessarily overlap with her participating in the performing of Swan Lake.

These final considerations bespeak a live avenue for future research into the precise philosophical assumptions of IIT, particularly insofar as the theory's formal tools are to be generalized and adopted for work within the various specialized sciences and the philosophical sub-disciplines pertaining to them, as suggested specifically for the philosophy

of social science and social ontology in this paper. Indeed, the *mereological* stance as to whether the parts of wholes are necessarily completely involved, it would seem, will make the difference as to whether IIT can be profitably utilized as a formal tool for the description of grades of ontological irreducibility.

Conclusion

In conclusion, recasting the individualism-holism debate in gradualist terminology seems a viable strategy for social ontology. This paper has argued that the precise formal tools needed for such an approach can be gained by introducing Integrated Information Theory (IIT) as a logical language for the analysis and discussion of our intuitions and claims about the ontological status of social entities. This utilization of IIT's formal tools has been shown to be principally feasible through their application to two toy models of social entities. The use of programming tools to apply IIT to social ontology is a notable advantage of the approach, since it potentially allows for the precise analysis and discussion of complex objects that are difficult to capture using traditional philosophical methods.

References

Bhargava, Rajeev. (1992). *Individualism in Social Science*, Clarendon Press.

Cerullo, Michael A. (2015). The Problem with Phi: A Critique of Integrated Information Theory. *PLoS Computational Biology*, *11* (9), 1-12.

Currie, Gregory. (1984). Individualism and Global Supervenience. *British Journal for the Philosophy of Science*, *35* (4), 345-358.

Engel, D., & Malone, T. W. (2018). Integrated information as a metric for group interaction. *PLOS ONE*, *13* (10), 1-19.

Epstein, B. (2009). Ontological individualism reconsidered. *Synthese*, *166* (1), 187-213.

Epstein, Brian. (2021). Social Ontology. *The Stanford Encyclopedia of Philosophy*, Edward N. Zalta (ed.), URL = https://plato.stanford.edu/archives/win2021/entries/social-ontology/.

Favela, Luis H. 2019. Integrated Information Theory as a Complexity Science Approach to Consciousness. In *Journal of Consciousness Studies*, *26* (1-2), 21-47.

Hoover, Kevin D. (1995). Is Macroeconomics for Real? *The Monist*, *78* (3), 235-257.

Holovatch, Y., Kenna, R., & Thurner, S. (2017). Complex systems: physics beyond physics. *European Journal of Physics*, *38* (2), 1-19.

Kincaid, Harold. (1986). Reduction, Explanation and Individualism. *Philosophy of Science*, *53* (4), 492-513.

Kincaid, Harold. (2014). Dead Ends and Live Issues in the Individualism-Holism Debate. In *Rethinking the Individualism-Holism Debate* (pp. 139-152). Springer Publishing.

Little, Daniel. (1991). *Varieties of Social Explanation*. Westview Press.

List, Christian & Kai Spiekermann. (2013). Methodological Individualism and Holism in Political Science: A Reconciliation. *American Political Science Review*, *107* (4), 629-643.

Marshall, W., Albantakis, L., & Tononi, G. (2018). Black-boxing and cause-effect power. *PLOS Computational Biology*, *14* (4), 1-21.

Mayner, W. G. P., Marshall, W., Albantakis, L., Findlay, G., Marchman, R., & Tononi, G. (2018). PyPhi: A toolbox for integrated information theory. *PLOS Computational Biology*, *14* (7), 1-21.

Mediano, P. A. M., Rosas, F. E., Farah, J. C., Shanahan, M., Bor, D., & Barrett, A. B. (2022).

Integrated information as a common signature of dynamical and information-processing complexity. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, *32* (1).

Mellor, D.H. (1982). The Reduction of Society. *Philosophy*, *57* (219), 51-75.

Oizumi, M., Albantakis, L., & Tononi, G. (2014). From the Phenomenology to the Mechanisms of Consciousness: Integrated Information Theory 3.0. *PLoS Computational Biology*, *10* (5), 1-25.

Pettit, Philip. (1993). *The Common Mind: An Essay on Psychology, Society, and Politics*. Oxford University Press.

Plato. (1997). The Sophist. In John M. Cooper & D. S. Hutchinson (Eds.) *The Complete Works* (pp. 234-294). Hackett Publishing Co.

Quinton, A. (1976). I—*The Presidential Address*: Social Objects. *Proceedings of the Aristotelian Society*, *76* (1), 1-28.

Sawyer, R. Keith. (2002). Nonreductive Individualism: Part I. *Philosophy of the Social Sciences*, *32* (4), 537-559.

Searle, J. R. (1995). *The Construction of Social Reality*. Free Press.

Stalnaker, Robert C. (1996). Varieties of Supervenience. *Philosophical Perspectives*, *10*, 221-241.

Tegmark, M. (2016). Improved Measures of Integrated Information. *PLOS Computational Biology*, *12* (11), 1-34.

Tononi, G. (2004). An information integration theory of consciousness. *BMC Neuroscience*, 5 (42).

Tononi, G. (2008) Consciousness as Integrated Information: a Provisional Manifesto. *The Biological Bulletin*, *215* (3), 216–242.

Tononi, G. (2012). Integrated information theory of consciousness: an updated account. *Archives Italiennes de Biologie*, *150*, 290-326.

Tononi, G., & Koch, C. (2015). Consciousness: here, there and everywhere? *Philosophical Transactions of the Royal Society B: Biological Sciences*, *370* (1668).

Tuomela, Raimo. (1989). Collective Action, Supervenience, and Constitution. *Synthese*, *80* (2): 243-266.

Wieland, Andreas. (2018). *The Supply Chain of a Computer*. Last accessed 20/04/2023 at: https://scmresearch.org/2018/09/28/the-supply-chain-of-a-computer/.

Zahle, J., & Collin, F. (2014). *Rethinking the Individualism-Holism Debate*. Springer Publishing.

Appendices

Appendix 1

```
In [17]: tpm = np.array([
    ...: [0,0,1],
    ...: [0,1,0],
    ...: [0,0,0],
    ...: [0,1,0],
    ...: [0,1,1],
    ...: [0,1,0],
    ...: [1,1,0],
    ...: [1,1,0]
    ...: ])

In [18]: cm = np.array([
    ...: [0,1,1],
    ...: [1,0,1],
    ...: [1,1,0]
    ...: ])

In [19]: network = pyphi.Network(tpm, cm=cm, node_labels=labels)

In [20]: subsystem = pyphi.Subsystem(network, state, node_indices
    ...: )

In [21]: state = (0,1,1)

In [22]: subsystem = pyphi.Subsystem(network, state, node_indices
    ...: )

In [23]: subsystem.cause_info((B, C), (A, B, C))
Out[23]: 1.5
```

Appendix 2

```
In [36]: tpm = np.array([
    ...: [0,0,1],
    ...: [0,1,0],
    ...: [0,0,0],
    ...: [0,1,0],
    ...: [0,1,1],
    ...: [0,1,0],
    ...: [1,1,0],
    ...: [1,1,0]
    ...: ])

In [37]: cm = np.array([
    ...: [0,1,1],
    ...: [1,0,1],
    ...: [1,1,0]
    ...: ])

In [38]: network = pyphi.Network(tpm, cm=cm, node_labels=labels)

In [39]: subsystem = pyphi.Subsystem(network, state, node_indices)

In [40]: state = (0,1,1)

In [41]: ces.labeled_mechanisms
Out[41]: (['A'], ['B'], ['C'], ['B', 'C'])
```

Appendix 3

```
In [29]: tpm = np.array([
    ...: [0,0,1],
    ...: [0,1,0],
    ...: [0,0,0],
    ...: [0,1,0],
    ...: [0,1,1],
    ...: [0,1,0],
    ...: [1,1,0],
    ...: [1,1,0]
    ...: ])

In [30]: cm = np.array([
    ...: [0,1,1],
    ...: [1,0,1],
    ...: [1,1,0]
    ...: ])

In [31]: network = pyphi.Network(tpm, cm=cm, node_labels=labels)

In [32]: subsystem = pyphi.Subsystem(network, state, node_indices)

In [33]: state = (0,1,1)

In [34]: sia = pyphi.compute.sia(subsystem)

In [35]: sia.cut
Out[35]: Cut [A, C] ──/ /──➤ [B]
```

Appendix 4

```
In [56]: cm = np.array([
    ...: [0,0,1],
    ...: [1,0,1],
    ...: [1,0,0]
    ...: ])

In [57]: subsystem = pyphi.Subsystem(network, state, node_indic
es)

In [58]: network = pyphi.Network(tpm, cm=cm, node_labels=labels
)

In [59]: subsystem = pyphi.Subsystem(network, state, node_indic
es)

In [60]: pyphi.compute.phi(subsystem)
Out[60]: 0.0

In [61]: ces = pyphi.compute.ces(subsystem)


In [62]: ces.labeled_mechanisms
Out[62]: (['A'], ['C'])
```

Appendix 5

```
In [43]: tpm = np.array([
    ...: [0,0,0],
    ...: [0,0,0],
    ...: [0,0,0],
    ...: [0,0,0],
    ...: [0,0,0],
    ...: [0,1,1],
    ...: [1,0,1],
    ...: [0,0,0]
    ...: ])

In [44]: cm = np.array([
    ...: [1,1,1],
    ...: [1,1,1],
    ...: [1,1,1]
    ...: ])

In [45]: network = pyphi.Network(tpm, cm=cm, node_labels=labels)

In [46]: state = (0,1,1)

In [47]: subsystem = pyphi.Subsystem(network, state, node_indices)

In [48]: subsystem.cause_effect_info((B, C), (A, B, C))
Out[48]: 0.75
```

Appendix 6

```
[In [169]]: subsystem.cause_effect_info((A,), (A,))
 Out[169]: 0.071429

[In [170]]: subsystem.cause_effect_info((A,), (A, B))
 Out[170]: 0.142856

[In [171]]: subsystem.cause_effect_info((A,), (A, C))
 Out[171]: 0.125

[In [172]]: subsystem.cause_effect_info((A,), (B, C))
 Out[172]: 0.125

[In [173]]: subsystem.cause_effect_info((A,), (A, B, C))
 Out[173]: 0.214284

[In [174]]: subsystem.cause_effect_info((A,), (A,))
 Out[174]: 0.071429

[In [175]]: subsystem.cause_effect_info((A,), (B,))
 Out[175]: 0.071429

[In [176]]: subsystem.cause_effect_info((A,), (C,))
 Out[176]: 0.0

[In [177]]: subsystem.cause_effect_info((A,), (A, B,))
 Out[177]: 0.142856

[In [178]]: subsystem.cause_effect_info((A,), (A, C,))
 Out[178]: 0.125

[In [179]]: subsystem.cause_effect_info((A,), (B, C,))
 Out[179]: 0.125

[In [180]]: subsystem.cause_effect_info((A,), (A, B, C,))
 Out[180]: 0.214284

[In [181]]: subsystem.cause_effect_info((B,), (A,))
 Out[181]: 0.125

[In [182]]: subsystem.cause_effect_info((B,), (B,))
 Out[182]: 0.125

[In [183]]: subsystem.cause_effect_info((B,), (C,))
 Out[183]: 0.0

[In [184]]: subsystem.cause_effect_info((B,), (A, B))
 Out[184]: 0.25

[In [185]]: subsystem.cause_effect_info((B,), (A, C))
 Out[185]: 0.125

[In [186]]: subsystem.cause_effect_info((B,), (B, C))
 Out[186]: 0.125

[In [187]]: subsystem.cause_effect_info((B,), (A, B, C))
 Out[187]: 0.25
```

(Appendix 6 continued)

```
In [188]: subsystem.cause_effect_info((C,), (A,))
Out[188]: 0.0

In [189]: subsystem.cause_effect_info((C,), (B,))
Out[189]: 0.0

In [190]: subsystem.cause_effect_info((C,), (C,))
Out[190]: 0.25

In [191]: subsystem.cause_effect_info((C,), (A, B))
Out[191]: 0.25

In [192]: subsystem.cause_effect_info((C,), (A, C))
Out[192]: 0.375

In [193]: subsystem.cause_effect_info((C,), (B, C))
Out[193]: 0.375

In [194]: subsystem.cause_effect_info((C,), (A, B, C))
Out[194]: 0.5

In [195]: subsystem.cause_effect_info((A, B), (A,))
Out[195]: 0.375

In [196]: subsystem.cause_effect_info((A, B), (B,))
Out[196]: 0.125

In [197]: subsystem.cause_effect_info((A, B), (C,))
Out[197]: 0.25

In [198]: subsystem.cause_effect_info((A, B), (A, B))
Out[198]: 0.5

In [199]: subsystem.cause_effect_info((A, B), (A, C))
Out[199]: 0.625

In [200]: subsystem.cause_effect_info((A, B), (B, C))
Out[200]: 0.375

In [201]: subsystem.cause_effect_info((A, B), (A, B, C))
Out[201]: 0.75

In [202]: subsystem.cause_effect_info((A, C), (A,))
Out[202]: 0.071429

In [203]: subsystem.cause_effect_info((A, C), (B,))
Out[203]: 0.071429

In [204]: subsystem.cause_effect_info((A, C), (C,))
Out[204]: 0.25

In [205]: subsystem.cause_effect_info((A, C), (A, B))
Out[205]: 0.5

In [206]: subsystem.cause_effect_info((A, C), (A, C))
Out[206]: 0.625
```

(Appendix 6 continued)

```
In [207]: subsystem.cause_effect_info((A, C), (B, C))
Out[207]: 0.375

In [208]: subsystem.cause_effect_info((A, C), (A, B, C))
Out[208]: 0.75

In [209]: subsystem.cause_effect_info((B, C), (A,))
Out[209]: 0.375

In [210]: subsystem.cause_effect_info((B, C), (B,))
Out[210]: 0.125

In [211]: subsystem.cause_effect_info((B, C), (C,))
Out[211]: 0.25

In [212]: subsystem.cause_effect_info((B, C), (A, B))
Out[212]: 0.5

In [213]: subsystem.cause_effect_info((B, C), (A, C))
Out[213]: 0.625

In [214]: subsystem.cause_effect_info((B, C), (B, C))
Out[214]: 0.375

In [215]: subsystem.cause_effect_info((B, C), (A, B, C))
Out[215]: 0.75

In [216]: subsystem.cause_effect_info((A, B, C), (A,))
Out[216]: 0.5

In [217]: subsystem.cause_effect_info((A, B, C), (B,))
Out[217]: 0.125

In [218]: subsystem.cause_effect_info((A, B, C), (C,))
Out[218]: 0.5

In [219]: subsystem.cause_effect_info((A, B, C), (A, B))
Out[219]: 1.0

In [220]: subsystem.cause_effect_info((A, B, C), (A, C))
Out[220]: 1.0

In [221]: subsystem.cause_effect_info((A, B, C), (B, C))
Out[221]: 0.875

In [222]: subsystem.cause_effect_info((A, B, C), (A, B, C))
Out[222]: 1.5
```

Appendix 7

```
In [61]: tpm = np.array([
    ...: [0,0,0],
    ...: [0,0,0],
    ...: [0,0,0],
    ...: [0,0,0],
    ...: [0,0,0],
    ...: [0,1,1],
    ...: [1,0,1],
    ...: [0,0,0]
    ...: ])

In [62]: cm = np.array([
    ...: [1,1,1],
    ...: [1,1,1],
    ...: [1,1,1]
    ...: ])

In [63]: network = pyphi.Network(tpm, cm=cm, node_labels=labels)

In [64]: state = (0,1,1)

In [65]: subsystem = pyphi.Subsystem(network, state, node_indices)

In [66]: ces = pyphi.compute.ces(subsystem)

In [67]: ces.labeled_mechanisms
Out[67]: (['A'], ['B'], ['C'], ['A', 'C'])

In [68]: ces.phis
Out[68]: [0.071429, 0.125, 0.25, 0.25]
```

Appendix 8

```
In [78]: tpm = np.array([
    ...: [0,0,0],
    ...: [0,0,0],
    ...: [0,0,0],
    ...: [0,0,0],
    ...: [0,0,0],
    ...: [0,1,1],
    ...: [1,0,1],
    ...: [0,0,0]
    ...: ])

In [79]: cm = np.array([
    ...: [1,1,1],
    ...: [1,1,1],
    ...: [1,1,1]
    ...: ])

In [80]: network = pyphi.Network(tpm, cm=cm, node_labels=labels)

In [81]: state = (0,1,1)

In [82]: subsystem = pyphi.Subsystem(network, state, node_indices)

In [83]: sia = pyphi.compute.sia(subsystem)

In [84]: sia.phi
Out[84]: 0.410218

In [85]: major_complex = pyphi.compute.major_complex(network, state)

In [86]: major_complex.subsystem.nodes
Out[86]: (A, B, C)
```

Appendix 9

```
In [507]: tpm = np.array([
     ...: [0,0,0,0,0,0],
     ...: [0,0,1,0,0,0],
     ...: [0,1,0,0,0,1],
     ...: [0,1,1,0,0,1],
     ...: [0,0,0,1,0,0],
     ...: [0,0,1,1,0,0],
     ...: [0,1,0,1,0,1],
     ...: [0,1,1,1,0,1],
     ...: [0,0,0,0,1,0],
     ...: [0,0,1,0,1,0],
     ...: [0,1,0,0,1,1],
     ...: [0,1,1,0,1,1],
     ...: [0,0,0,1,1,0],
     ...: [0,0,1,1,1,0],
     ...: [0,1,0,1,1,1],
     ...: [0,1,1,1,1,1],
     ...: [0,0,0,0,0,0],
     ...: [0,0,1,0,0,0],
     ...: [0,1,0,0,0,1],
     ...: [0,1,1,0,0,1],
     ...: [0,0,0,1,0,0],
     ...: [0,0,1,1,0,0],
     ...: [0,1,0,1,0,1],
     ...: [0,1,1,1,0,1],
     ...: [0,0,0,0,1,0],
     ...: [0,0,1,0,1,0],
     ...: [0,1,0,0,1,1],
     ...: [0,1,1,0,1,1],
     ...: [0,0,0,1,1,0],
     ...: [0,0,1,1,1,0],
     ...: [0,1,0,1,1,1],
     ...: [0,1,1,1,1,1],
     ...: [0,0,0,0,0,0],
     ...: [0,0,1,0,0,0],
     ...: [0,1,0,0,0,1],
     ...: [0,1,1,0,0,1],
     ...: [0,0,0,1,0,0],
     ...: [0,0,1,1,0,0],
     ...: [0,1,0,1,0,1],
     ...: [0,1,1,1,0,1],
     ...: [0,0,0,0,1,0],
     ...: [0,0,1,0,1,0],
     ...: [0,1,0,0,1,1],
     ...: [0,1,1,0,1,1],
     ...: [0,0,0,1,1,0],
     ...: [0,0,1,1,1,0],
     ...: [0,1,0,1,1,1],
     ...: [0,1,1,1,1,1],
     ...: [0,0,0,0,0,0],
     ...: [0,0,1,0,0,0],
     ...: [0,1,0,0,0,1],
     ...: [0,1,1,0,0,1],
     ...: [0,0,0,1,0,0],
     ...: [0,0,1,1,0,0],
     ...: [0,1,0,1,0,1],
     ...: [0,1,1,1,0,1],
     ...: [1,0,0,0,1,0],
     ...: [1,0,1,0,1,0],
     ...: [1,1,0,0,1,1],
     ...: [1,1,1,0,1,1],
     ...: [1,0,0,1,1,0],
     ...: [1,0,1,1,1,0],
     ...: [1,1,0,1,1,1],
     ...: [1,1,1,1,1,1]
     ...: ])
```

(Appendix 9 continued)

```
In [517]: cm = np.array([
     ...: [0,0,1,0,0,0],
     ...: [0,1,0,0,0,1],
     ...: [0,0,0,1,0,0],
     ...: [0,0,0,0,1,0],
     ...: [1,0,0,0,0,0],
     ...: [1,0,0,0,0,0]
     ...: ])

In [518]: labels = ('CM', 'RM', 'CL',
 'C', 'CO', 'SC')

In [519]: network = pyphi.Network(tpm
, cm=cm, node_labels=label
     ...: s)

In [520]: state = (1,1,1,1,1,1)

In [521]: subsystem = pyphi.Subsystem
(network, state, node_indi
     ...: ces)

In [522]: CM, RM, CL, C, CO, SC = sub
system.node_indices

In [523]: sia = pyphi.compute.sia(sub
system)

In [524]: sia.phi
Out[524]: 0.0
```