

Is the Integrated Information Theory of Consciousness Compatible with Russellian Panpsychism?

Hedda Hassel Mørch

DRAFT – under review

Abstract: The Integrated Information Theory (IIT) is a leading theory of consciousness in neuroscience, which implies a kind of panpsychism. In this paper, I consider whether IIT is compatible with Russellian panpsychism, a particular kind of panpsychism which arguably avoids the main problems of both physicalism and dualism. I will first show that if IIT were compatible with Russellian panpsychism, it would contribute to solving Russellian panpsychism’s combination problem, which threatens to show that the view does not solve the main problems of physicalism and dualism after all. I then show that the theories are not compatible as they currently stand, in view of what I call the coarse-graining problem. I will explain this problem, and offer two possible solutions involving small modifications of IIT. Given either of these modifications, IIT and Russellian panpsychism may be fully compatible after all, and jointly enable significant progress on the mind-body problem.

1 Introduction

Panpsychism is the view that every physical thing is associated with consciousness. More precisely, it is the view that every physical thing is either (1) conscious as a whole, (2) made of parts which are all conscious, or (3) itself forms part of a greater conscious whole. Humans and animals (or certain areas of human and animal brains) are conscious in the first sense – our consciousness is unified, or has a single, subjective point of view. Panpsychists typically think fundamental particles, such as quarks, are conscious in this sense too. Artifacts and arbitrary aggregates, such as tables, chairs and heaps, on the other hand, may only be conscious in the second sense, in virtue of being made of particles that are conscious in the first sense. Particles and neurons that form part of human and animal brain areas which are conscious in the first sense are perhaps only conscious in the third sense. By “consciousness”, panpsychists mean

phenomenal consciousness, i.e., the property of having some form of subjective, qualitative experience, or of there being *something that it is like for* something to exist.¹

Panpsychism has recently received new attention in both philosophy and neuroscience. In philosophy of mind, it has been argued that a version of panpsychism, panpsychist Russellian monism (named after Bertrand Russell, for reasons that will be explained below), is able to avoid the most serious problems that confront the traditional main positions in the field, physicalism and dualism (Nagel 1979; Strawson 2006a; Alter and Nagasawa 2012; Chalmers 2015). For physicalism, this is the problem of the epistemic gap – aspects of which are highlighted by the knowledge argument (Jackson 1982, 1986) and the conceivability argument (Kripke 1980; Chalmers 1996). For dualism, it is the problem of mental causation (Kim 1988; Papineau 2001).

In neuroscience, the Integrated Information Theory of consciousness (IIT) (Tononi 2004, 2008, 2011; Tononi, Albantakis, and Oizumi 2014) is a new fundamental theory of consciousness which has proved to have significant explanatory and predictive power, and is currently among the leading theories within the field. According to IIT, consciousness is correlated with *integrated information*, also referred to as Φ . *Information* in IIT measures, roughly, the extent to which a system causally constrains its own immediate past and future state.² *Integration* measures, roughly, the extent to which this information depends on the causal interconnectivity

¹ Consciousness does not require the capacity for, e.g., thought, abstraction or self-awareness. These can be regarded as advanced, complex forms of consciousness and reserved for advanced, complex physical systems. Therefore, panpsychism does not entail that particles think or are self-conscious.

² I.e., how much can you know about the previous state of the system only by looking at its present state, given that all external influences are fixed (given that you know all the relevant causal laws)? And how much can you know about the next state of the system, fixing external influences? The measure also takes into account the number of possible states a system can be in – the greater the number of possible states of a system, the more past and future states can be ruled out by the current state, which in IIT corresponds to higher information.

of the system's parts.³ Since “even simple matter has a modicum of Φ ” (Koch 2012: 132), IIT implies panpsychism in the sense defined above.⁴

IIT sometimes encounters the objection that although it may appear to account well for empirical data about human consciousness – such as the absence of consciousness during deep sleep, general anesthesia, epileptic seizures, and in regions of the brain such as the cerebellum, all of which correlate with low values of Φ (Tononi and Koch 2015; Tononi et al. 2016) – it can't be correct because panpsychism is absurd (Searle 2013). But if Russellian panpsychism solves the main philosophical problems of consciousness, IIT's panpsychism should be

³ I.e., how much information would you lose by dividing the system in half, and (unidirectionally) cutting all connections from one half to the other? Computers, of the kind we make today, might in principle have the same quantity of information as human brains – they could have a similar number of possible states and causally constrain their own past and future states to a similar degree. But in today's computers, the connectivity between transistors is limited and mainly feed-forward (i.e., one transistor gives input to a second transistor, which then gives input to a third transistor, but not the other way around). Cutting connections (unidirectionally) from the output level to the input level of a feed-forward system is not going to lead to any loss of information. In the brain, in contrast, especially in (parts of) the cerebrum, there is significant feedback connectivity – neurons mutually constrain a vast number of other neurons. Cutting connections between two parts of the cerebrum in any direction is therefore going to lead to a great loss of information. Hence, the cerebrum has a much higher level of *integrated* information (Φ). For a more detailed introduction to IIT, see Tononi and Koch (2015). For a full description of the latest version of the theory, see Tononi, Albantakis, and Oizumi (2014).

⁴ Tononi also acknowledges that IIT comes close to panpsychism (Tononi 2008: 236; Tononi and Koch 2015: 11). He is careful to distinguish IIT's panpsychism from the universalist form of panpsychism (to be discussed further below) according to which all things, including artifacts and aggregates, are conscious as individuals, i.e., in the first sense. But IIT still implies that all things are conscious in at least the second or third sense. As Koch notes, “protons and neutrons consist of a triad of quarks that are never observed in isolation. They constitute an infinitesimal integrated system” (Koch 2012: 132). If protons and neutrons have some Φ , then these particles will be conscious either individually, i.e., in the first sense, or as parts of systems with greater Φ than themselves (such as atoms or molecules), i.e., in the third sense. If protons and neutrons are conscious, then all things that are made of protons and neutrons will at least be conscious as groups, i.e., in the second sense. And if quarks are always part of integrated systems like protons and neutrons, they will necessarily be conscious in the third sense.

One might wonder about other fundamental particles, such as electrons, that unlike quarks may be found in isolation. Would they be conscious? This question does not have a straightforward answer. As a simplification, IIT defines Φ only for systems that can be regarded as composed of discrete elements. Physics is now widely taken to support the view that particles are really excitations in continuous fields, hence IIT does not directly apply to fundamental physics. Barrett (2014) has given one proposal for how IIT can be extended to fundamental physics, which explicitly entails that isolated electrons will have some Φ , and therefore some consciousness (2014: 4), roughly because field excitations are complexly structured events. However, it is not guaranteed that other ways of extending IIT to physics will have the same result. Therefore, IIT can only be said to imply (complete) panpsychism in the sense of strongly suggesting it, it does not strictly entail it. In view of this, the central question of this paper may be more precisely put as: “Is IIT, if extended to *fundamental physics in some way that entails (complete) panpsychism*, compatible with Russellian panpsychism?”.

regarded as a feature, not a bug. IIT could claim philosophical support in virtue of being the only current neuroscientific theory compatible with the solution to the mind-body problem.

But not everyone agrees that Russellian panpsychism solves the mind-body problem. The main objection to Russellian panpsychism is based on the so-called combination problem (James 1890; Seager 1995).⁵ The combination problem consists in explaining how complex consciousness, of the kind we know from our own case, could result from the appropriate combination of entities with simple consciousness, i.e., the particles that constitute our brain and according to panpsychism possess rudimentary forms of consciousness. According to the objection, the combination problem gives rise to further problems which are strongly analogous the main problems of physicalism and dualism (Goff 2006, 2009; Chalmers 2015, 2016).

At first glance, it appears IIT can help Russellian panpsychism solve the combination problem. The precise claim of IIT is that a system is conscious, in the first sense, if and only if it is a *maximum* of Φ , i.e., if the system has higher Φ than any of its own parts as well as any bigger system of which it forms part. This is IIT's *Exclusion* postulate – consciousness does not overlap. This means that whenever the Φ of a system surpasses the Φ of the parts, the system goes from being conscious in the second, aggregative sense, to being conscious in the first, unified sense. This explains how mental combination occurs, at least in certain respects. Tononi and Koch imply (2015: 11, 13) that IIT may offer a complete solution to the combination problem in and of itself. But one might also regard IIT as merely a contributing factor to a complete solution, when taken together with other proposed solutions to the problem. Other proposed solutions include the phenomenal bonding view (Goff 2009, 2016) and the fusion view (Seager 2010, 2016; Mørch 2014). These proposals look promising in purely metaphysical respects, but both arguably fail to show what could possibly constitute an

⁵ The most *common* objection to panpsychism is probably “the incredulous stare”. The most *serious* objection is the combination problem.

empirical correlate of combination. IIT shows what this correlate could be, namely maximal Φ .

Unfortunately, however, a solution to the combination problem in terms of IIT is not straightforward. In order to solve the combination problem, IIT needs to be compatible with *Russellian* panpsychism, not just panpsychism in general. Panpsychism could also be combined with dualist or physicalist metaphysics,⁶ but will then have no advantage compared to standard physicalism or dualism with respect to the mind-body problem. And there are some indications that IIT is only compatible with dualist or physicalist panpsychism. This would be bad news for both IIT and Russellian panpsychism: IIT could no longer claim that its panpsychism is in fact a philosophical advantage, and Russellian panpsychism would retain its combination problem.

In this paper, I will explain the indications of this disappointing result, and argue that it mainly derives from what I will call *the coarse-graining problem*. I will show that this is not a surface problem: IIT and Russellian panpsychism are in fact incompatible as the theories currently stand. But the conflict could still be resolved with the introduction of some substantive new principles. I will suggest two candidates for such principles: first, a modification of IIT's *Exclusion* postulate (not abolition, as some have called for (Schwitzgebel 2015)); second, a modification of IIT's principle for determining the spatio-temporal grain of conscious systems. My arguments, if successful, will not support a fully satisfactory solution to the combination problem in terms of IIT, but they should give reason for optimism and motivate further research into this approach.

⁶ Dualist panpsychism would be the view that every physical thing is connected with a mental substance (or wholly non-physical mental properties) via fundamental psychophysical laws of nature. Physicalist panpsychism would be the view that consciousness is to be reductively identified with a ubiquitous physical property (such as integrated information understood in a purely functionalist way).

I will begin by explaining the metaphysics of Russellian panpsychism and the combination problem in more detail. Then I will discuss how IIT could help with the combination problem, either on its own, or combined with the phenomenal bonding or fusion view. Then I will present the coarse-graining problem, which shows that the views are in fact incompatible, followed by my two proposals for how to resolve the problem.

2 Physicalism, Dualism and Russellian Panpsychism

According to Russellian panpsychism, consciousness is associated with every physical thing because consciousness constitutes the intrinsic nature of physical relational structure. As Bertrand Russell noted (1927), physics (or the physical sciences) describes only relational or structural properties – how physical things relate to other things, but not what they are like in and of themselves. These are the kind of properties which can be described in the austere language of mathematics. The phenomenal properties of consciousness, in contrast, seem to be intrinsic – we know what they are like in and of themselves, independently of how they relate to other things. According to Russellian panpsychism, structural properties need non-structural realizers, and relations need relata with intrinsic properties. Furthermore, phenomenal properties are the only intrinsic properties we know. Russellian panpsychists infer from this that consciousness could be the realizer of all physical structure, or an intrinsic property of every entity described by physics in relational terms (Seager 2006; Alter and Nagasawa 2012).⁷

As noted, the main attraction of Russellian panpsychism is that it offers a solution to the mind-body problem which appears to avoid the main difficulties of physicalism and dualism at once.

⁷ Note that Russell himself ultimately concluded that the intrinsic character of physical structure is not mental but neutral, i.e., neither mental nor physical. The more general view that the intrinsic character of physical structure is either neutral, protomental or mental is known as Russellian monism. Despite its name, central elements of Russellian monism predate Russell. Notable predecessors include Leibniz (see Pereboom 2015), Schopenhauer (see especially 1966a, pp. 119-127 and 1966b, pp. 191-200), James (1890, 1912), Mach (1886, 1894, 1905) and arguably Kant (especially given the reading defended in Langton 1998). Furthermore, the contemporary version of the view may not fully correspond to Russell's own neutral monism (Stubenberg 2015; but see Wishon 2015 for a different interpretation).

Physicalism is the view that phenomenal (i.e., conscious) properties are identical to, or constituted by, physical properties – where physical properties are understood as the kind of properties that physics (or the physical sciences) describes. Its main problem is that it appears impossible in principle to see any entailment from the existence of physical properties to the existence of any phenomenal properties – something which should be possible if the latter were indeed identical to or constituted by the former (Chalmers 2003) (and we have an adequate grasp of the nature of both the mental and the physical, see Goff (2016)). In the absence of such an entailment, we are able to conceive of the mental and the physical coming apart, as the conceivability argument (i.e., zombie argument) highlights (Chalmers 2009). Also, someone who knows all physical properties would not be able to deduce the existence and nature of mental properties, as the knowledge argument highlights (Jackson 1982).

Dualism is the view that phenomenal properties are distinct from physical properties, and are causally generated or affected by them. Its main problem is the threat of epiphenomenalism, because there is no non-redundant causal role for non-physical properties if the physical world is causally closed. The principle of physical causal closure is the principle that every physical event (that has a cause) has a sufficient physical cause. This has solid empirical and methodological support (Papineau 2001). It seems implausible that some physical events are systematically overdetermined by both mental and physical causes and, in any case, overdetermination still renders the mental redundant. The final option for dualists who accept causal closure and deny overdeterminism is to deny the causal efficacy of the mental and embrace epiphenomenalism.

Russellian panpsychism avoids physicalism's main problem because it does not say that phenomenal properties are identical with or constituted by physical properties. Instead, there is a sense in which phenomenal properties constitute physical properties, as their intrinsic realizers. If physical properties are structural properties, then there is in fact an entailment from

the existence of phenomenal, intrinsic properties to the existence of physical properties – if some entities with intrinsic properties exist, then they necessarily stand in *some* relations and form *some* structure (though we can't say a priori *which* structure).⁸

Russellian panpsychism avoids dualism's main problem because this inverted constitution relation does not result in epiphenomenalism. If phenomenal properties realize physical structure, they are essential to the physical world. Causally efficacious physical properties can't even exist without realizers or intrinsic properties. At the same time, this hypothesis leaves physical structure unchanged, so the physical world remains causally closed from the structural point of view. Russellian panpsychism can therefore respect the scientific considerations that support physical causal closure.

3 The Combination Problem

The combination problem is the problem of explaining how complex consciousness, of the kind we know from our own case, can result from the combination of simple entities with simple consciousness. *Constitutive panpsychism* is the view that complex consciousness is constituted by simple entities with simple consciousness, or, in other words, that complex consciousness is nothing over and above a collection of simple conscious experiences causally and/or spatiotemporally related in a certain way. This view faces a problem analogous to physicalism's problem of the epistemic gap: it appears impossible in principle to see any entailment from the existence of entities with simple micro-consciousness – in any purely spatiotemporal or causal relation – to the existence of complex macro-consciousness. This can be illustrated by panpsychist versions of both the conceivability argument and the knowledge

⁸ The entailment also goes the other way around. If we accept that structure needs intrinsic realizers, and that phenomenal properties are the only intrinsic properties we know, then there is an entailment from the existence of physical properties to the existence of either phenomenal properties or unknown properties of a similar nature (non-panpsychist Russellian monists prefer the latter alternative).

argument – think of microphenomenal zombies (Goff 2009) or a physically and micro-phenomenally omniscient Mary (Chalmers 2016).

Emergent panpsychism is the view that complex consciousness is distinct from micro-consciousness causally and/or spatio-temporally related in a certain way, but is rather causally generated by micro-conscious configurations. This view faces a problem of macro-mental causation analogous to dualism’s problem of mental causation: macro-consciousness appears causally and explanatorily redundant⁹ assuming that micro-consciousness already realizes all physical structure. Positing additional structure for which distinct macro-consciousness could be the realizer would bring emergent panpsychism into conflict with the principle of physical (or at least *microphysical*) causal closure.

In this way, the combination problem threatens to undermine the motivation for Russellian panpsychism. It appears to show that the view does not avoid the problems of physicalism and dualism after all; it merely moves them, and has them reappear in a different guise.

4 IIT and the Combination Problem

IIT is a theory with both empirical and metaphysical aspects. On the one hand, it proposes an empirical correlate of consciousness: all and only conscious systems are maxima of Φ . On the other hand, it aims to explain *why* this correlation holds. Tononi argues that the correlation follows *a priori* from a set of phenomenological axioms, truths about consciousness which are self-evident upon introspection (Tononi, Albantakis, and Oizumi 2014).

The correlational claim, that maximum Φ is the correlate of consciousness, is empirically verifiable in principle. Although Φ can be precisely calculated for complex systems only in

⁹ Perhaps outright excluded: overdeterminism is perhaps not even an option for emergent panpsychism, because one structural property can’t have two distinct mental realizers (microphenomenal properties plus an emergent macrophenomenal property).

theory; insofar as it can be estimated, it predicts which kinds of alterations in the brain would lead to a loss of consciousness (namely those that reduce the Φ of a brain area below the Φ of its parts). It is also verifiable in terms of how well it accounts for pre-existing data about consciousness. IIT accounts for why an area of the brain such as the cerebellum is not conscious, even though it has more neurons than the cerebrum; why epileptic seizures, as well as deep sleep, lead to loss of consciousness even though neural activity remains high or increases; and a number of other facts that puzzle neuroscientists.¹⁰ Other proposed correlates of consciousness, such as activation in specific brain areas, electrical activity within some particular frequency range, or the P300 wave, all seem to be absent in some clearly conscious states (Tononi and Koch 2015: 3-4).

The explanatory claim, that the correlation follows *a priori* from phenomenological axioms, is mainly supported by appeal to philosophical argument. If the explanatory claim is correct, then IIT could potentially solve the combination problem on its own. The *a priori* connection would eliminate all epistemic gaps, and the *Exclusion* postulate eliminates the problem of causal exclusion (this will be discussed in more detail below). The philosophical case for the explanatory claim is, however, somewhat contentious.¹¹ It is worth considering, therefore, whether the empirically verifiable, purely correlational claim could support a solution to the combination problem without commitment to the explanatory claim. I will suggest two ways in which the correlational claim can be combined with metaphysical approaches to combination that have been developed independently of IIT.

¹⁰ The cerebellum has low interconnectivity between neurons in spite of having a large number of them (Tononi 2008: 221); under epileptic seizures and deep sleep Φ strongly decreases even though activity increases or remains high (Tononi 2008: 223).

¹¹ Some complain that the alleged phenomenological axioms are less than self-evident (e.g., Cerullo 2015). Some argue that the empirical postulates are not adequately supported (e.g., Schwitzgebel 2015: section 2, concerning *Exclusion*). There might also be multiple ways of translating the empirical postulates into a mathematical formula (Tegmark 2016). Chalmers (2016) claims that IIT is best seen as providing a fundamental, *a posteriori* law of nature specifying when combination occurs, but he does not specify precisely how he thinks Tononi's *a priori* justification fails.

The phenomenal bonding view (Goff 2016) is a proposal for a solution to the combination problem for constitutive panpsychism. According to this view, some physical (i.e., spatial, temporal and/or causal) relations have their own intrinsic nature which is not reducible to the intrinsic properties of their relata. Physics is silent about the intrinsic nature of these relations, just as it is silent about the intrinsic nature of physical relata. Via introspection, we have access to the intrinsic nature of some physical relata (i.e., the phenomenal properties of our own brains), but we don't have access to the intrinsic nature of physical relations. According to the phenomenal bonding view, if we knew the intrinsic nature of the kind of relation that connects the particles in our brain, we would see that the existence of macro-consciousness is indeed entailed by the existence of micro-conscious entities thus related.

One of the main challenges for the phenomenal bonding view is to specify what physical relation the phenomenal bonding relation could correspond to (Chalmers 2016). The phenomenal bonding relation must be a physical relation which holds within all and only systems where it is plausible that mental combination takes place. Goff suggests that the spatial relation is the phenomenal bonding relation. This would entail that all phenomenal properties which are spatially related are also co-conscious – experienced together from a single subjective point of view – so that every thinkable object (group of particles) is conscious as a whole (in the first sense). Many panpsychists (including Strawson (2006a: 26) and Tononi and Koch (2015: 11) reject this kind of universalism – they want to restrict unified consciousness to just some kinds of systems, including brains and other organs (or organisms), and perhaps also lower level natural objects like minerals or molecules, and/or certain kinds of computers or robots, etc.¹² There is no obvious candidate for a physical relation which gives this result.

¹² What is the argument against universalism about combination? The view is clearly highly counterintuitive, but so is panpsychism itself (at least for most people within Western cultural traditions), so this should not count for too much. One more substantive worry with universalism is that it is very hard to imagine what it could be like to be an aggregate such as, say, the combination my toe and a piece of my shoe, an arbitrary 30% area of

IIT tells us what the phenomenal bonding relation could be. If maximal Φ is the correlate of unified consciousness, then the phenomenal bonding relation would be equivalent to a certain kind of causal relation: causal relations that relate elements within systems that have more Φ than overlapping systems. One might wonder *why* the intrinsic nature of causal relations would depend on whether or not they are part of systems with maximal Φ .¹³ But the most important thing for the phenomenal bonding view is that there exists some natural, not wildly disjunctive, kind of physical relation, that does not entail universalism, and could possibly be identified as the phenomenal bonding relation.

The fusion view (Seager 2010, 2016; Mørch 2014) is an approach to the combination problem for emergent panpsychism. According to this view, the emergence of macro-consciousness is not synchronic – emergent macro-consciousness does not exist on top of, and at the same time as, an underlying micro-conscious base. Rather, the emergence is diachronic – emergent macro-consciousness succeeds and replaces the set of microphenomenal properties that gives rise to it. The generation of macro-consciousness happens as fusion: under certain conditions, microphenomenal properties fuse or blend together to form a new macro-conscious unity. The

the brain area that supports my consciousness, or the set of two people fighting. One might retort that it is also very hard to imagine what it is like to be a fundamental particle. But it is a different challenge to imagine what it is like to be a simple, but unified thing, than to imagine what it would be like to be a wildly disunified thing. Phenomenological simplicity seems hard to imagine mainly simply because our power of imagination is limited, whereas phenomenological disunity or incoherence seems hard to imagine because it is hard to even make sense of. The experience of a system such as <my toe and a piece of my shoe> or <an arbitrary 30% area of the brain area that supports my consciousness>, and <the set of two people fighting> would have to be a very incoherent and meaningless experience, perhaps even paradoxical. In the case of <the set of two people fighting> it would involve being conscious of contradictory thoughts (person 1 is thinking “person 2 is wrong”, person 2 is thinking “person 2 is not wrong”; the aggregate is having both thoughts). In the case of <30% of a brain area> it would seem to involve something like thinking 30% of a thought, or having 30% of a feeling, which is not clearly possible. There is nothing similarly borderline paradoxical about the idea of a very simple but still unified experience. This objection is clearly not definitive – there are probably ways for universalists to respond. But for those who worry about whether there could be a fully satisfactory response to this and other objections to universalism, IIT gives a way of avoiding it all together.

¹³ Chalmers notes (2016: section 6.3) that relations along the lines of IIT could be the phenomenal bonding relation, but objects that these relations are derivative and not fundamental. He claims that a solution to the combination problem in terms of phenomenal bonding and something like IIT get a new combination problem: how do fundamental causal relations (proto-bonding relations) become phenomenal bonding relations when they take on IIT-like character? A phenomenal bonding theorist could perhaps respond by renewed appeal to our ignorance of the intrinsic nature of physical relations: if we knew the intrinsic nature of causal relations, we would see how it works.

fusion view avoids the problem of macro-mental causation: if macro-consciousness replaces micro-consciousness (i.e., micro-consciousness *transforms* into macro-consciousness), it is the sole candidate realizer for the physical properties with which it is correlated. It is hard to see how a synchronic emergence view can get around the same problem.

The main challenge for the fusion view is to find a physical correlate of mental fusion. Seager (2010, 2016) claims that quantum entanglement and the formation of black holes are physical examples of fusion. In both entangled systems and black holes, the constituents lose their individuality and are supplanted by an irreducible new unity – a “large simple”. But nothing like this seems to be going on in the brain. Mørch (2014) has argued that mental fusions don’t need to be partless simples, they only need to be wholes which are metaphysically *prior* to their parts. But no parts of the brain clearly appear to be prior to their parts either.

IIT is in itself a kind of fusion view. According to the *Exclusion* postulate, only maxima of Φ are conscious, hence, whenever a complex system becomes conscious by gaining maximal Φ , its parts will lose micro-consciousness (in the first sense; but they will remain conscious in the third sense, as defined in the introduction). Φ measures not only information but also integration, and integration is a form of unity.¹⁴ IIT thus defines a sense in which physical systems can be more unified than their parts, and thereby be said to be metaphysically prior to their parts, as fusions should be. One might wonder whether fusions shouldn’t ideally be characterized by even stronger marks of unity.¹⁵ But the most important thing for the fusion view is that there is some property which could serve to identify fusions from the physical point of view, which is empirically detectable, but does not entail any revisionary physical

¹⁴ Recall, high integration means that the parts of a system are highly causally interconnected, and thereby bound together as a unit.

¹⁵ Tononi claims that maximal Φ corresponds to maximal causal power. This would give fusions a very strong form of metaphysical priority, without strong emergence of the kind that violates microphysical causal closure. But the view that Φ corresponds with causal power depends on a distinctive and controversial metaphysical view about causation (as described in Hoel, Albantakis, and Tononi 2013).

hypotheses (such as quantum entanglement at the brain level, or other kinds of strong physical emergence).

5 The Coarse-Graining Problem

Unfortunately, there is a problem which gets in the way of any solution to the combination problem in terms of IIT: the coarse-graining problem. This problem stems from the way IIT identifies the spatiotemporal grain of consciousness.

IIT not only selects a set of elements which regarded as a system gives maximal Φ . It also selects a spatiotemporal grain (Hoel, Albantakis, and Tononi 2013; Hoel et al. 2016). One might regard the brain (or brain area that correlates with consciousness) as a system consisting of elements at the smallest spatial grain – as a system of subatomic particles at the nanometer scale or below. Calculating on this basis will result in certain value Φ . But one might also regard the brain as a system consisting of elements at a coarser spatial grain – such as a system of neurons or groups of neurons. This will result in a higher value of Φ (Hoel et al. 2016: 11-12). One can also vary the temporal grain. One can regard the system as proceeding from state to state at the rate of a nanosecond, and calculate how much integrated information the present nanosecond-long state of system has about the immediate past and future nanosecond. Or one can regard each state as lasting for milliseconds, and calculate how much integrated information the present milliseconds-long state has about the immediate past and future milliseconds long states. For the brain, a coarser temporal grain (perhaps corresponding to the time it takes for neurons to fire) will result in higher Φ .

If a maximally integrated system exists at a certain spatiotemporal grain, the structure below that grain will not matter to the quality of the experience of that system (Tononi and Koch 2015: footnote 9). For this reason, IIT allows for multiple microphysical realizability of systems with identical consciousness. If the Φ of the brain peaks when the brain is considered

as a system of neurons, it won't matter to the character of its experience whether the microstructure of the neurons is carbon-based or silicon-based as long as they function identically at the neuron level.

This leads to a conflict with Russellian panpsychism, as follows. Dispositions are causal properties involving relations between circumstances (or inputs) and manifestations (or outputs), such as “will repel electrons if nearby”, “will attract massive particles in proportion to their distance”, or “will travel at the speed of light if unobstructed”. Every major theory of causation entails that the dispositions of things are determined by their intrinsic properties together with the laws of nature. Or more precisely, they entail that dispositions are thus determined given that things have any intrinsic properties in the first place and that these properties are not epiphenomenal.¹⁶ According to the causal powers view (Shoemaker 1980; Ellis 2001; Bird 2007), dispositions are determined by the intrinsic, or intrinsically grounded, causal powers of individual things (these powers in turn determine the laws).¹⁷ According to realism about laws (Dretske 1977; Tooley 1977; Armstrong 1978), dispositions are determined by irreducible governing laws, and it is natural to think that the laws apply to things in virtue of their intrinsic properties.¹⁸ According to the regularity theory (Hume 1739-40; Lewis 1973), the dispositions of things depend on their intrinsic, qualitative properties

¹⁶ Theories such as ontic structural realism (Ladyman and Ross 2007) deny that intrinsic properties exist, claiming that dispositions understood as purely structural properties are absolutely fundamental, i.e., do not supervene on anything. It also seems coherent to suppose that dispositional properties are ungrounded in the way ontic structural realists claim, while claiming that intrinsic properties such as phenomenal properties exist but are epiphenomenal.

¹⁷ Some proponents of the causal powers view hold that powers are absolutely fundamental and have no categorical (i.e., intrinsic and non-dispositional) aspects or grounds. Russellian panpsychism, which takes phenomenal properties to be the categorical grounds of dispositions, is incompatible with this version of the view. However, other proponents of the view claim that powers have a categorical aspect (Martin and Heil 1999; Strawson 2008). Russellian panpsychists could say the categorical aspects of properties are phenomenal and ground the dispositional aspects.

¹⁸ Armstrong explicitly holds that laws of nature consist in relations between categorical, i.e., intrinsic, universals.

and regularities involving properties of the same type across the actual or all possible worlds (the laws are regarded as reducible to these regularities).

According to Russellian panpsychism, the intrinsic properties of things are all phenomenal properties,¹⁹ and phenomenal properties are not epiphenomenal. It follows that if Russellian panpsychism is true, dispositions at least *nomologically* supervene on phenomenal properties. That is to say, given that the laws of nature (whether irreducible, derived from intrinsic powers, or reducible to contingent regularities) remain the same, then how things are disposed to behave is determined by what they are intrinsically like, i.e., their phenomenal properties.

If dispositions nomologically supervene on phenomenal properties, then there can't be things with the same phenomenal properties but different dispositions, given the same laws. But according to IIT's coarse-graining principle, there *can* be things with the same phenomenal properties but different dispositions given the same laws, namely coarse-grained systems such as silicon and organic brains: at the coarse-grained level, they are functionally identical, but at the fine-grained, microphysical level, they have different dispositions corresponding to their different silicon and carbon constituents. It follows that Russellian panpsychism is false – some dispositions do not nomologically supervene on phenomenal properties.²⁰

¹⁹ This is not to say that *all* intrinsic properties are phenomenal, according to Russellian panpsychism, only the intrinsic properties of individual things. As discussed above, Goff's phenomenal bonding view says that some relations between individual things, namely spatial relations, have their own, non-phenomenal intrinsic properties responsible for combination. If IIT is combined with the phenomenal bonding view, in the way I proposed above, then causal relations also involve this kind of intrinsic properties.

²⁰ Note that Russellian panpsychism does not require that all physical properties nomologically supervene on phenomenal properties, only that dispositional properties do. Physical structure as a whole also consists of how dispositions manifest, and the circumstances that trigger their manifestation, such as spatial relations. These relations might be primitive, or perhaps supervene on non-phenomenal, intrinsic properties of relations (such as those involved in phenomenal bonding). If all physical properties had to nomologically supervene on phenomenal properties, Russellian panpsychism would come closer to idealism than many consider it to be.

Given this, one might think the differences between silicon and carbon brains come down to differences in non-phenomenal circumstances, not differences in dispositions. However, at the fundamental level, silicon and carbon consists of different numbers and ratios of fundamental particles such as quarks, electrons and so on. Each type of fundamental particle has its own fundamental and unique dispositions. Therefore, silicon and carbon brains must be regarded as having different sets of fundamental dispositions.

One might think this conflict can simply be resolved by saying that although silicon brains and organic brains have identical macrophenomenal properties, they have different microphenomenal properties. Silicon neurons (or molecules/atoms) would have silicon-type micro-experience, while the organic neurons (or molecules/carbon atoms) would have different carbon-type micro-experience, while both constituting the same human macro-experience. The identical human macro-experiences nomologically determine identical coarse-grained macrophysical dispositions, but the different neuron (or molecule/atom) experiences will nomologically determine the different microphysical dispositions.

But this is ruled out by IIT's *Exclusion* postulate, according to which consciousness never overlaps – only the system with maximal Φ is conscious. Neurons (or molecules/atoms) can't have different experiences if they are excluded from having experiences of their own in the first place. Hence, IIT is not compatible with Russellian panpsychism in view of the coarse-graining principle and the *Exclusion* postulate.

It might seem, then, that the best way of rendering IIT and Russellian panpsychism compatible would be to abandon one of these principles. But abandoning either leads to both philosophical and empirical problems.

6 Abandoning *Exclusion*

Rejecting *Exclusion* leads to a *prima facie* very undesirable proliferation of consciousness. Nations, the internet, the galaxy and the universe all have non-zero, but non-maximal, Φ and would therefore be conscious without *Exclusion*. There would also be multiple overlapping consciousnesses within the brain. Not only would each neuron be conscious, there would also be one consciousness associated with almost any combination of neurons (the brain, the brain minus one neuron, minus two neurons, minus two other neurons, and so on). In other words, we get a serious case of the *problem of the many*. For those who are motivated to combine IIT

with Russellian panpsychism in order to avoid universalism with respect to phenomenal bonding, this consequence should be unacceptable.

Rejecting *Exclusion* is also empirically problematic for IIT. As mentioned, the empirical case for IIT is largely based on data about when our (waking or dreaming) consciousness disappears and reappears according to reports, such as in deep sleep, general anesthesia, seizures and coma. These states don't leave the brain with zero Φ – our consciousness disappears (i.e., disintegrates into multiple lesser consciousnesses) long before Φ reaches zero. What explains this? Officially, IIT is not entirely clear on this point. One possible explanation is that low, non-maximal but non-zero, Φ still correlates with some dim form of consciousness, but not with memory or reportability.²¹ But this seems somewhat stipulative – why would it necessarily be the case that low Φ states can't be remembered or reported? It could also be very worrisome if true about anesthesia!²² An alternative, more systematic explanation would be to appeal to the *Exclusion* postulate: our consciousness disappears because the Φ of the normally conscious brain area goes below the Φ of neurons, neuron groups, or other subsystems that constitute it. Given *Exclusion*, these subsystems will become individually conscious instead (each to a much lesser degree) because they are now the maxima. It is not clear whether the Φ of sleeping or anesthetized brain areas is actually lower than the Φ of any subsystems – given the practical difficulties with measuring Φ precisely, this is still an open empirical question. But by abandoning *Exclusion*, this simple and elegant explanation will be ruled out regardless of the empirical outcome. In view of these problems, abandoning *Exclusion* in order to render IIT compatible with Russellian panpsychism looks like it won't be worth the costs.

²¹ Tononi and Koch (2015: 13) gesture toward such a view.

²² One might think that consciousness during anesthesia is not painful, or if it is, that it is just very dimly painful and therefore nothing to worry about – but none of this is obvious.

7 Abandoning Coarse-Graining

Instead of *Exclusion*, one could consider abandoning coarse-graining. This would mean that when calculating the Φ of a system, one should always regard it as a system of fundamental particles. If so, silicon and organic brains would have differences in their experiences on which the differences in their microphysical dispositions could nomologically supervene.

But abandoning coarse-graining also leads to both philosophical and empirical problems. Firstly, the view that our experience is coarse-grained is not only, or mainly, motivated by IIT. It is widely regarded as a phenomenological datum. When we look at our own experience, it directly appears, phenomenologically, to have a certain spatiotemporal structure, and this structure appears to be much more coarse-grained than the microphysical structure of our brain. As already noted, we don't appear to experience as many distinctions (or other sort of grains) within our experience as there are particles (or other sorts of grains) in our brain. Furthermore, our experiences appear to last much longer than the shortest microphysical events in the brain. Experiences appear to last for at least some milliseconds, but microphysical events in the brain last nanoseconds or less.

This is the basis for the so-called *grain problem* in philosophy of mind known from Lockwood (1993) and Sellars (1965, 1971): the grain of experience does not correspond to the microphysical grain of the brain. The conclusion often drawn from the grain problem is that human consciousness needs an irreducibly *macrophysical* correlate, which is precisely what it gets with IIT's coarse-graining principle, but loses without it.

Secondly, rejecting coarse-graining gives an empirical problem. As noted, if we regard the brain as a system of particles, it will have lower Φ than if we regard it as a system of neurons. And it will probably also have lower Φ than small neuron groups, individual neurons or perhaps even molecules or atoms in the brain. Therefore, without coarse-graining, we might have to

say that small neuron groups, neurons, molecules or atoms in the brain are conscious, not the brain (or any larger brain area) as a whole. This would result in at least thousands (perhaps billions) other consciousnesses in our brains, each of which as highly conscious as our own. This seems very implausible (as another version of the problem of the many, although in this case the many do not spatially overlap) and does not fit any natural interpretation of empirical data. With coarse-graining, in contrast, Φ will peak only in a few locations, and the highest peak would be much higher than the others, so there will only be one most advanced consciousness in the brain.²³

Hence, abandoning coarse-graining seems out of the question for IIT. What I will now suggest is that, in order to render IIT compatible with Russellian panpsychism, there is no need to wholly abandon either the coarse-graining or the *Exclusion* postulate. Instead, it would suffice to merely modify either one of the principles, in ways I will now describe.

8 Modifying *Exclusion*

One way of avoiding the coarse-graining problem is to modify the *Exclusion* postulate by relativizing it to spatiotemporal grains. Instead of saying that consciousness never overlaps, IIT could say that (1) consciousness never overlaps *within the same spatiotemporal grain*, and (2) there is consciousness only at grains with higher Φ than any grain below. This would entail that the brain, as a coarse-grained system, will have overlapping consciousness at the finest, microphysical grain (as systems at the finest grain vacuously satisfy clause (2), since there is no grain below). In two brains with different fine-grained constituents, such as silicon and carbon, the overlapping microconsciousness will also be different. Differences in

²³ At least in normal circumstances – according to IIT, split brain (severed corpus callosum) patients will have one consciousness in each hemisphere, with almost the same Φ . It is also possible that our consciousness temporarily splits into two in cases such as automatic (absent-minded) driving.

microphysical dispositions could then nomologically supervene on these differences in microconsciousness.

This relativized *Exclusion* postulate avoids the problems that result from wholly abandoning it. It will not prevent human consciousness from disappearing in deep sleep, general anesthesia and so on before Φ goes to zero. Human consciousness would disappear when the Φ of its correlate brain area goes lower than the Φ of individual neurons or smaller neuron groups, because *Exclusion* is still in effect within this grain. For the same reason, there will not be overlapping consciousness in overlapping systems of neurons (the brain minus one neuron, the brain minus two neurons, two other neurons, etc.). It will not result in internet, galaxy or cosmic consciousness, because the brain has maximal Φ all grains considered. The internet, the galaxy or the cosmos will not have higher Φ than the brain if we regard them as systems made of neurons and other entities at the same grain as the brain, nor if we regard them as systems made of coarser-grained parts such as computers or stars, or finer-grained parts such as atoms. At smaller scales, there could be other grains between the subatomic and neuron level that increase Φ , for example systems of molecules, resulting in more than one level of microconsciousness, but there will not be arbitrarily or absurdly many such levels.

This modification of *Exclusion* is only compatible with *constitutive* Russellian panpsychism. If microphenomenal and macrophenomenal properties are entirely distinct, as per the emergent view, then microphenomenal properties at lower grains would causally exclude macrophenomenal properties at higher levels with maximal Φ . But the idea that coarse-grained macrophenomenal properties are constituted by overlapping fine-grained microphenomenal properties leads to another problem.

Say I have an experience of a tiny patch or “pixel” of phenomenal green, the smallest patch of green that is possible given the grain of my experience. This green patch will be completely

uniform green for me, because my experience can't contain any complexity below its own smallest grain. Now, this uniform green patch will actually be constituted by microphenomenal properties at a smaller grain, for example even smaller patches of blue and yellow that are the experiences of the particles in my brain (or some other kinds of particle experiences that we can't imagine). But then it seems my experience of green is both uniform and partless, and complex, made of distinct blue and yellow (or otherwise non-green) parts, at the same time.

Many Russellian panpsychists, and non-reductionists about consciousness in general, are motivated by the principle that there is no appearance/reality-distinction for phenomenal properties (if an experience *appears* to be of uniform green, then it *is* of uniform green, phenomenally speaking). Given this principle, it would not be possible for the same phenomenal property to appear either uniform green or complex yellow/blue (or otherwise non-green) depending on the grain. If the same property appears in all of these ways, then the same property would also *be* both uniform and non-uniform, which is contradictory. Hence, this way of rendering IIT and Russellian panpsychism seems to come at the cost of having to introduce a kind of appearance/reality distinction for phenomenal properties which may be in tension with Russellian panpsychism's initial motivation.

Note, however, that the same kind of problem would confront constitutive panpsychism independently of this proposal in view of the grain problem alone. The grain problem, discussed above, also claims – on a general phenomenological basis as opposed to IIT's more specific and theoretical one²⁴ – that some macrophenomenal properties appear uniform. But according to constitutive panpsychism, all macrophenomenal properties are constituted by microphenomenal properties and hence really complex. This problem is a version of what is

²⁴ As discussed above, IIT's coarse graining principle may be partially justified on the basis of the grain problem, but another motivation is an independent theoretical one according to which coarse-graining is required for Φ to increase in the right places in the brain.

known as the revelation problem, and can be regarded as a further aspect of the combination problem for constitutive panpsychism (Goff 2006; Chalmers 2016).

Therefore, this is not a problem that results from this proposal in particular. It might still, however, be regarded as a limitation of the proposal, because it means that there is an important aspect of the combination problem that this particular way of combining IIT with Russellian panpsychism does nothing to solve. The revelation problem may still be resolvable in other ways, of course, though as of yet this has arguably not been clearly demonstrated.²⁵

9 Modifying Coarse-Graining

Another way of avoiding the coarse-graining principle, that also avoids the problems associated with overlapping uniform and complex experiences, is to rather modify the coarse-graining principle – in a way that starts from noticing a certain tension between two aspects of the combination problem.

The grain problem, as already discussed, says that macro-consciousness has *too little structure* to reflect the full microphysical structure of our brains. But there is also another aspect of the combination problem, known as *the palette problem* (Chalmers 2016: section 7). According to this problem, macro-consciousness has *too many qualities*. In physics, we find a limited number of fundamental particles (about 17, according to the standard model). This suggests a correspondingly limited number of basic microphenomenal qualities. In our experience, however, we find what appears to be an endless number of different phenomenal qualities (colors, sounds, emotions and so on). It is hard to see how all these qualities can result (without

²⁵ Proposed solutions to the revelation problem include Strawson (2006b: 252), Chalmers (2016: 190) and Brogaard (2016: 147-150). For criticism of some of these proposals, see Goff (2015) on Strawson and Brogaard (2016) on Chalmers. It is also worth noting that Russell seemed to take our acquaintance with phenomenal properties to be compatible with a certain kind of appearance/reality distinction, as illustrated by phenomenal continua cases in which different phenomenal properties can be falsely judged to be identical (Wishon 2017) (thanks to an anonymous referee for pointing this out). One might think this view can also account for appearance/reality distinctions of the kind implied by this proposal.

radical emergence) just from combining a small number of basic microphenomenal qualities in different ways.

If macro-consciousness has too little structure, but too many qualities, to reflect its microphysical correlate, this suggests that the missing structure is somehow *encoded* in the extra qualities – that microphysical structure is reflected in macrophenomenal *qualities* as opposed to macrophenomenal *structure*. Instead of having one little distinct patch in our experience corresponding to every particle in the brain, perhaps a single coarse-grained quality (like a patch of color, or an emotion) reflects complex microphysical structure in virtue of the many complex ways in which it is different from other possible macrophenomenal qualities.

Accordingly, IIT could be modified to say that fine-grained microphysical structure matters to the quality, but not the structure of experiences. Or in other words, that macrophenomenal structure will be multiply microphysically realizable,²⁶ but macrophenomenal qualities will not be. If your organic neurons were to be replaced by functionally identical silicon neurons, then your experience would have the same (contrastive and motivational) structure, but different qualities (i.e., a shifted spectrum). For example, for the silicon brain, tomatoes and grass will be experienced as having color-like qualities we cannot imagine, but they will be as different from each other as red and green are for us, and they will have pain and pleasure-like qualities which feel different from ours, but they still feel equally bad and good. The fact that silicon and organic brains have the same macrophenomenal structure could account for why these systems will have the same dispositions at the macrolevel. That they have different, shifted qualitative spectra could perhaps account for why they have different microphysical

²⁶ According to Russellian panpsychism, there is a sense in which the mental realizes the physical. But there is also realization within the physical, in another sense of the term. To say that macroconsciousness is multiply realizable given Russellian panpsychism would really be to say that the macrophysical *correlate* of macroconsciousness is multiply realizable physically speaking, while macroconsciousness realizes the physical correlate metaphysically speaking.

dispositions (why they respond differently to microscopic measurement and decompose in different ways).

The distinction between quality and structure need not be sharp. As long as there are some aspects of coarse-grained experiences that fall neatly into the qualitative category, others may belong to both. One might think that coarse-grained physical structure correlates with the structural aspects of the content of coarse-grained experiences, and those qualitative aspects that *are not* separable from the structural aspects (those aspects for which a shifted spectrum *is not* possible). Fine-grained physical structure correlates with those qualitative aspects that *are* separable from phenomenal structure (those aspects for which a shifted spectrum *is* possible).

Many questions remain as to precisely how microphysical causal structure could be encoded or grounded in unstructured qualities, as well as how this could be incorporated into the formalism of IIT.²⁷ I have only shown that there is some extra qualitative information in one place that can be connected to missing structural information in another place, and there is nothing that immediately rules out that there is some way of translating one to the other. Given this, it at least does not look impossible for IIT and Russellian panpsychism to be combined.

10 Conclusion

I have suggested two ways of resolving the coarse-graining problem and rendering IIT and Russellian panpsychism compatible. These suggestions involve substantive modifications of some basic principles of IIT, either the *Exclusion* postulate or the coarse-graining principle. Given one of these modifications, IIT would support (significant progress towards) a solution

²⁷ Note that IIT as it stands makes no distinction between structural and qualitative aspects of the content of consciousness. IIT distinguishes between the overall level, or amount, of consciousness (which corresponds to the overall Φ of a system) and the content of consciousness (which corresponds to the precise causal structure of a system, in IIT called conceptual structure). My proposal is to make a further distinction, within content, between structural and qualitative aspects. This does not involve any change to the way coarse-graining affects overall Φ .

to the combination problem for Russellian panpsychism – either on its own, in view of the explanatory claim according to which the principles of mental combination are *a priori* deducible from phenomenological axioms, or on the basis of its correlational claim only, in combination with either the phenomenal bonding view or the fusion view of mental combination.

Both my suggested modifications leave open further questions. The suggested modification of the *Exclusion* postulate is in tension with the claim that phenomenal qualities necessarily *are* as they appear, and thereby fails to resolve the aspect of the combination problem known as the revelation problem. The suggested modification of the coarse-graining principle assumes mysterious links between qualities and structure. In the absence of a clear solution to the revelation problem, perhaps the mystery of the latter proposal is preferable to the tension of the former. Modifying the *Exclusion* postulate might also be more at odds with Tononi’s own philosophical case for IIT.

IIT has many philosophical aspects which have not been discussed here – as noted, I have set the theory’s central philosophical claims mostly aside and mainly considered the significance of its in principle empirically verifiable correlational claim, i.e., the claim that maximal Φ is the correlate of consciousness. IIT as a whole should be subjected to closer philosophical scrutiny if the theory is to fully support a solution to the combination problem. On the one hand, one might find even more sources of tension with Russellian metaphysics within IIT. On the other hand, one might discover other aspects of IIT that could contribute to even further advancing the case for a panpsychist solution to the mind-body problem.²⁸

²⁸ Thanks to Many thanks to Giulio Tononi, Erik Hoel, David Chalmers, Torin Alter, Luke Roelofs, Neil Mehta, Kelvin McQueen and participants and referees at the conferences “IIT: Foundational Issues” (NYU 2015), “Non-physicalist Views of Consciousness” (Cambridge University 2016), “The Science of Consciousness” (Tucson, Arizona 2016) and “Minds Online” (2016) for helpful comments on this paper.

Bibliography

- Alter, Torin, and Yujin Nagasawa. 2015. *Consciousness in the Physical World: Perspectives on Russellian Monism*. Oxford University Press.
- Alter, Torin, and Yujin Nagasawa. 2012. What Is Russellian Monism? *Journal of Consciousness Studies* 19 (9-10): 67-95.
- Armstrong, David M. 1978. *A Theory of Universals. Universals and Scientific Realism Volume II*. Cambridge University Press.
- Barrett, Adam B. 2014. An Integration of Integrated Information Theory with Fundamental Physics. *Frontiers in Psychology* 5: 63.
- Brogaard, Berit. 2016. In Search of Mentons. In *Panpsychism: Contemporary Perspectives*, eds. G. Brüntrup and L. Jaskolla. Oxford: Oxford University Press.
- Bird, Alexander. 2007. *Nature's Metaphysics: Laws and Properties*. Oxford: Clarendon Press.
- Cerullo, Michael A. 2015. The Problem with Phi: A Critique of Integrated Information Theory. *PLOS Computational Biology* 11 (9).
- Chalmers, David J. 1996. *The Conscious Mind: In Search of a Fundamental Theory*. New York: Oxford University Press.
- . 2003. Consciousness and Its Place in Nature. In *Blackwell Guide to Philosophy of Mind*, eds. S. P. Stich and T. A. Warfield. Malden, MA: Blackwell.
- . 2009. The Two-Dimensional Argument against Materialism. In *Oxford Handbook to the Philosophy of Mind*, eds. B. P. McLaughlin and S. Walter. Oxford: Oxford University Press.
- . 2015. Panpsychism and Panprotopsychism. In *Consciousness in the Physical World: Essays on Russellian Monism*, eds. T. Alter and Y. Nagasawa. New York: Oxford University Press.
- . 2016. The Combination Problem for Panpsychism. In *Panpsychism: Contemporary Perspectives*, eds. G. Brüntrup and L. Jaskolla. Oxford: Oxford University Press.
- Dretske, Fred I. 1977. Laws of Nature. *Philosophy of Science* 44 (2): 248-268.
- Ellis, B. D. 2001. *Scientific Essentialism*. Vol. 113. Vol. 450 Cambridge University Press.
- Goff, Philip. 2006. Experiences Don't Sum. *Journal of Consciousness Studies* 13 (10-11): 53-61.
- . 2009. Why Panpsychism Doesn't Help Us Explain Consciousness. *Dialectica* 63 (3): 289-311.
- . 2015. Against Constitutive Russellian Monism. In *Consciousness in the Physical World: Perspectives on Russellian Monism*, eds. T. Alter and Y. Nagasawa Oxford University Press.
- . 2016. The Phenomenal Bonding Solution to the Combination Problem. In *Panpsychism: Contemporary Perspectives*, eds. G. Brüntrup and L. Jaskolla. Oxford: Oxford University Press.
- Hoel, Erik P., Larissa Albantakis, William Marshall, and Giulio Tononi. 2016. Can the Macro Beat the Micro? Integrated Information across Spatiotemporal Scales. *Neuroscience of Consciousness* 2016 (1).
- Hoel, Erik P., Larissa Albantakis, and Giulio Tononi. 2013. Quantifying Causal Emergence Shows That Macro Can Beat Micro. *Proceedings of the National Academy of Sciences* 110 (49): 19790-19795.
- Hume, David. 1739-40. *A Treatise of Human Nature*.
- Jackson, Frank. 1982. Epiphenomenal Qualia. *Philosophical Quarterly* 32 (April): 127-136.
- . 1986. What Mary Didn't Know. *The Journal of Philosophy* 83 (5): 291-295.
- James, William. 1890. *The Principles of Psychology*. Vol. 1. London: Macmillan.

- . 1912. *Essays in Radical Empiricism*. New York, London, Bombay and Calcutta: Longmans, Green & Co.
- Kim, Jaegwon. 1988. Explanatory Realism, Causal Realism, and Explanatory Exclusion. *Midwest Studies in Philosophy* 12 (1): 225-239.
- Koch, Cristof. 2012. *Consciousness: Confessions of a Romantic Reductionist*. Cambridge, MA: MIT Press.
- Kripke, Saul A. 1980. *Naming and Necessity*. Cambridge, MA: Harvard University Press.
- Ladyman, James, and Don Ross. 2007. *Every Thing Must Go: Metaphysics Naturalized*. Oxford: Clarendon Press.
- Langton, Rae. 1998. *Kantian Humility: Our Ignorance of Things in Themselves*. Oxford: Oxford University Press.
- Lewis, David. 1973. Causation. *Journal of Philosophy* 70 (17): 556-567.
- Lockwood, Michael. 1993. The Grain Problem. In *Objections to Physicalism*, ed. H. M. Robinson Oxford University Press.
- Mach, Ernst. 1886. *Beiträge Zur Analyse Der Empfindungen*. Jena: Verlag von Gustav Fischer.
- . 1894. *Populärwissenschaftliche Vorlesungen*. Reprint of the fifth edition, Vienna: Böhlau Verlag, 1923.
- . 1905. *Erkenntnis Und Irrtum*. Fifth edition translated as *Knowledge and Error*, Dordrecht: Reidel, 1976.
- Marshall, William, Larissa Albantakis, and Giulio Tononi. 2016. Black-Boxing and Cause-Effect Power. *ArXiv e-prints* 608.03461.
- Martin, C. B., and John Heil. 1999. The Ontological Turn. *Midwest Studies in Philosophy* 23 (1): 34-60.
- Mørch, Hedda Hassel. 2014. Panpsychism and Causation: A New Argument and a Solution to the Combination Problem (Doctoral Dissertation), Departement of Philosophy, Classics, History of Art and Ideas, University of Oslo, Oslo.
- Nagel, Thomas. 1979. Panpsychism. In *Mortal Questions*. Cambridge: Cambridge University Press.
- Nagasawa, Yujin, and Torin Alter. 2015. *Consciousness in the Physical World: Perspectives on Russellian Monism*. Oxford University Press.
- Papineau, David. 2001. The Rise of Physicalism. In *Physicalism and Its Discontents*, eds. C. Gillett and B. Loewer. Cambridge: Cambridge University Press.
- Pereboom, Derk. 2015. Consciousness, Physicalism, and Absolutely Intrinsic Properties. In *Consciousness in the Physical World: Perspectives on Russellian Monism*, eds. T. Alter and Y. Nagasawa Oxford University Press.
- Russell, Bertrand. 1927. *The Analysis of Matter*. London: Kegan Paul, Trench, Trubner & Co.
- Schopenhauer, Arthur. 1966a. *The World as Will and Representation* Vol. 1. Translated by E. F. J. Payne. New York: Dover.
- Schopenhauer, Arthur. 1966b. *The World as Will and Representation* Vol. 2. Translated by E. F. J. Payne. New York: Dover.
- Schwitzgebel, Eric. 2015. If Materialism Is True, the United States Is Probably Conscious. *Philosophical Studies* 172 (7): 1697-1721.
- Seager, William. 1995. Consciousness, Information and Panpsychism. *Journal of Consciousness Studies* 2 (3): 272-288.
- . 2006. The 'Intrinsic Nature' Argument for Panpsychism. *Journal of Consciousness Studies* 13 (10-11): 129-145.
- . 2010. Panpsychism, Aggregation and Combinatorial Infusion. *Mind and Matter* 8 (2): 167-184.

- . 2016. Panpsychist Infusion. In *Panpsychism: Contemporary Readings*, eds. G. Brüntrup and L. Jaskolla. Oxford: Oxford University Press.
- Searle, John R. 2013. Can Information Theory Explain Consciousness? *New York Review of Books* 60 (1).
- Sellars, Wilfrid S. 1965. The Identity Approach to the Mind-Body Problem. *Review of Metaphysics* 18 (March): 430-51.
- . 1971. Seeing, Sense Impressions, and Sensa: A Reply to Cornman. *Review of Metaphysics* 24 (March): 391-447.
- Shoemaker, Sydney. 1980. Causality and Properties. In *Time and Cause: Essays Presented to Richard Taylor*, ed. P. van Inwagen. Dordrecht: Reidel.
- Strawson, Galen. 2006a. Realistic Monism: Why Physicalism Entails Panpsychism. *Journal of Consciousness Studies* 13 (10-11): 3-31.
- Strawson, Galen. 2006b. Panpsychism? Reply to Commentators with a Celebration of Descartes. *Journal of Consciousness Studies* 13 (10-11): 184-280.
- . 2008. The Identity of the Categorical and the Dispositional. *Analysis* 68 (4): 271-282.
- Stubenberg, Leopold. 2015. Russell, Russellian Monism and Panpsychism. In *Consciousness in the Physical World: Perspectives on Russellian Monism*, eds. T. Alter and Y. Nagasawa Oxford University Press.
- Tegmark, Max. 2016. Improved Measures of Integrated Information. *PLOS Computational Biology* 12 (11).
- Tononi, Giulio. 2004. An Information Integration Theory of Consciousness. *BMC Neuroscience* 5 (1): 42.
- . 2008. Consciousness as Integrated Information: A Provisional Manifesto. *The Biological Bulletin* 215 (3): 216-242.
- . 2011. Integrated Information Theory of Consciousness: An Updated Account. *Archives Italiennes de Biologie* 150 (2-3): 56-90.
- Tononi, Giulio, Larissa Albantakis, and Masafumi Oizumi. 2014. From the Phenomenology to the Mechanisms of Consciousness: Integrated Information Theory 3.0. *PLOS Computational Biology* 10 (5).
- Tononi, Giulio, Melanie Boly, Marcello Massimini, and Christof Koch. 2016. Integrated Information Theory: From Consciousness to Its Physical Substrate. *Nature Reviews Neuroscience* 17 (7): 450-461.
- Tononi, Giulio, and Christof Koch. 2015. Consciousness: Here, There and Everywhere? *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 370 (1668).
- Tooley, Michael. 1977. The Nature of Laws. *Canadian Journal of Philosophy* 7 (4): 667-98.
- Wishon, Donovan. 2015. Russell on Russellian Monism. In *Consciousness in the Physical World: Perspectives on Russellian Monism*, eds. T. Alter and Y. Nagasawa Oxford University Press.
- Wishon, Donovan. 2017. Russellian Acquaintance and Frege's Puzzle. *Mind* 126 (502): 321-370.