# What it might be like to be a group agent

Max F. Kramer
University of Arizona

This is a post-peer-review, pre-copyedit version of an article published in *Neuroethics*. The final authenticated version is or will soon be available online at: http://dx.doi.org/10.1007/s12152-021-09459-7. Please cite final version.

Abstract: Many theorists have defended the claim that collective entities can attain genuine agential status. If collectives can be agents, this opens up a further question: can they be conscious? That is, is there something that it is like to be them? Eric Schwitzgebel [1] argues that yes, collective entities (including the United States, taken as a whole), may well be significantly conscious. Others, including Kammerer [2], Tononi and Koch [3], and List [4] reject the claim. List does so on the basis of Tononi's Integrated Information Theory of consciousness [5]. I argue here that List's rejection is too quick, and that groups can, at least in principle, display the kind of informational integration we might think is necessary for consciousness. However, group consciousness will likely differ substantially from the individual experiences that give rise to it. This requires the defender of group consciousness to face up to a similar combination problem as the panpsychist.

Keywords: social groups, consciousness, Integrated Information Theory, the Combination Problem, collective intentionality

## §1. Introduction

Most philosophers admit that there can be collective entities, with holdout only from some extreme individualists [6, §3.1.2]. Some even believe that such entities have an existence separate from the existence of their individual members [7–9]. Though there are some who have reservations [10], many philosophers believe that certain groups have the proper internal organization to reach agential status [11]. To that end, it has been claimed that groups can have emotional states of their own [12]. My concern in this paper is a similar matter, though one that may be even more outlandish from the perspective of common sense. I will defend the view that collective entities can, in principle, be conscious, and that we have some reason to think that extant collective entities are capable of achieving consciousness. More specifically, I will argue against Christian List's [4] claim that Integrated Information Theory rules out the possibility of (significant) collective consciousness. Before I plunge

in, there are two matters to straighten out: first, why would we care about whether there can be collective consciousness? And second, what is consciousness, anyway? I'll take these in reverse order.

We use the term 'conscious' to mean multiple things. A distinction in usages owing to Block [13] enjoys widespread endorsement, though how the distinction ought to be drawn is not always a point of agreement. One way we might understand consciousness is in the sense of awareness, e.g., 'I am conscious of the fact that it is my wife's birthday.'[1] This is either called awareness or, in Blockean terms, access consciousness. Then, there is the 'what it is like' of subjective experience—the redness of an object in the visual field, the sound of a C# on a Steinway, etc. This is phenomenal consciousness.[2] For ease of keeping these two notions separate, I will use the term 'awareness' to refer to access consciousness and 'consciousness' to refer to phenomenal consciousness.

Following List, I am interested in the claim that collective entities can be conscious, which is the more counterintuitive claim and the one that has received more attention in the past few years.[3] The idea that groups could be aware might at first be counterintuitive as well, but some examples are usually enough to sway intuitions. For instance, we often talk of corporations being aware of economic conditions in making business decisions. To the extent that we can talk about any sort of collective computational thought in a non-metaphorical sense, even if such thought is fully determined by the thoughts of individuals within the collective, we are admitting that the group can be aware. The functional equivalent of awareness is usually a component of arguments for group agency. Collective consciousness, on the other hand—the subjective experience of a collective entity—is much more difficult to wrap one's mind around. Thus, corporate agents[4] are seen as the equivalent of philosophical zombies [18]—entities that can perform seemingly intelligent behavior but without any subjective experience—albeit quite strange and, perhaps, dumb ones, depending on the range of behaviors of which the collective is capable.

Another reason (aside from glee in provocation) to defend the consciousness of groups is the possibility that consciousness is required for an important form of responsibility known as accountability. This claim was recently defended by Baddorf [19] along the following lines: (1) If collectives are not capable of consciousness, then collectives do not have minds; (2) If collectives do

[1] Those who defend a robust cognitive phenomenology might take issue with my characterization, because they hold that there is a 'something it is like' to be aware of some fact. For clarification, I mean awareness to be something along the lines of 'potentially accessible in working memory.'
[2] Though see [14] for a rejection of this distinction.
[3] For some other treatments, which I will largely set aside for the purposes of this paper, see [15–17].
[4] Throughout, I will use the terms 'group,' 'corporate,' 'collective,' and 'plural' interchangeably.

not have minds, then they cannot be accountable. (3) Collectives are not capable of consciousness; therefore, (4) Collectives cannot be accountable.[5] I will take Baddorf's argument (except for (3)) at face value. Thus, those philosophers who wish to defend the notion that collectives can be irreducibly responsible in all the ways other agents can be must attempt to show that there is something that it is like to be a collective entity.

## §2. Schwitzgebel's conditional argument for collective consciousness

Unfortunately, Baddorf has little to say to defend the claim that collective entities, in principle, cannot be conscious, aside from appeal to intuitions and the fact that we would need to change our moral practices if this were the case. The second is not an argument for the in-principle claim—we can coherently hold that collectives are *in principle* capable of being conscious and that no collective has ever met (and even that no collective is ever *likely* to meet) the standards for consciousness. Making an in-principle claim on the basis of mere intuitions seems to be begging the question.

More to the point, there is an extant argument for the (conditional) consciousness of collective entities. Schwitzgebel [1] claims that materialist theories of consciousness—that is, theories that hold that subjective experience is a product of matter being organized in a certain pattern—gives us good reason to endorse the claim that at least some group entities are conscious. I'll run through the argument quickly. First, materialists do not deny consciousness to non-human animals. For example, most materialists think that rabbits enjoy conscious experience, and that consciousness is not limited to human brains. Moreover, if materialists endorse the claim that rabbits are conscious, they are also disposed to accept the claim that naturally-evolved aliens, even if they are physiologically very different than we are, can be conscious if they are behaviorally very similar to us. In other words, consciousness depends not on manifesting any particular organization of matter but on whether that matter is organized in such a way that it produces a host of functions we identify as consciousness. This is a broadly functionalist account of consciousness, which is the species of materialist account that currently enjoys the most favor in the philosophical and scientific literature on consciousness. This is not to say that brains are not central to theories of consciousness—to the contrary, any live account of consciousness must deal with the physiology and architecture of the human brain. Schwitzgebel is explicit on this point:

[5] Similarly, Rupert [10] argues that having phenomenal states is a requirement for intentionality, and therefore that groups lack mental states in virtue of lacking phenomenal states.

> According to materialism, what's really special about us is our brains. Brains are what make us conscious…Now what is so special about brains, on the materialist view? Why do they give rise to conscious experience while a similar mix of chemical elements in chicken soup does not? *It must be something about how those elements are organized* [1, p. 1706; emphasis mine].

This is, presumably, why we think that rabbits are conscious. They have something (their brain) that performs for them the relevant set of functions that our brain performs for us. Moreover, we can see why this is the case when we perform functional neuroimaging studies on rabbit brains. They are organized in a similar way to how ours are, and that organization breeds somewhat similar results, as evidenced by their behavior. At the same time, this functionalist paradigm allows us to sensibly imagine all manner of alien creatures with different physiological organizations but in whom those organizations output similar functionality to ours. In fact, we can imagine conscious aliens whose cognitive architecture is distributed throughout their bodies but is still as informationally integrated as our brain is, such that they report (linguistically!) a single unified stream of subjective experience.

If we admit that such a functional organization is, in principle, capable of producing consciousness, then we should also admit that were these aliens able to detach portions of their body while remaining functionally integrated, they would continue to enjoy the same level of consciousness. At this point, there is no reason to hold on to the morphological prejudice that consciousness requires a spatially contiguous functional organization. Thus, any collective entity that was organized in a way that achieved an appropriate functional organization could be a subject of conscious experience. To make his point as vividly as possible, Schwitzgebel further argues that "the United States, conceived of as a spatially distributed, concrete entity with people as some or all of its parts, literally possesses a stream of conscious experience over and above the experiences of its members considered individually" [1, p. 1698].

It is natural to balk at this suggestion. It is one thing (the balking goes) for an entity's cognitive architecture to be spatially distributed and entirely another for the spatially distributed cognitive architecture to be composed of objects *that are themselves conscious*, as in the case of the United States. After all, we could quibble about whether neurons in the human brain are spatially contiguous or distributed at a microphysical level, but there seems to be a major difference in kind between a neuron and a person, each taken as one node in an organized network. In a version of this objection, some philosophers have defended what Schwitzgebel terms 'anti-nesting principles.' This includes Putnam

[20], who, in a discussion of pain, stipulates that "no organism capable of feeling pain possesses a decomposition into parts which are separately capable of feeling pain" [1, p. 1702]. All that appears to underwrite this principle is a desire to avoid supposedly counterintuitive results, such as the possibility that a swarm of bees could feel pain as a single entity. This patently begs the question against the possibility of collective consciousness.

Tononi [21], at least, gives grounds for his anti-nesting principle: the 'exclusion postulate.' On his account, consciousness is a matter of integrating information. When one informationally-integrated network is nested within another, consciousness only occurs at the level that integrates the most information. He defends the principle by appeal to Occam's razor, and I take him to mean that we ought not proliferate consciousness beyond where we are justified in thinking it is present.

This argument runs in step with another kind of exclusion: causal exclusion. Kim [22, 23] famously argued for a principle of causal exclusion, which rejected the possibility of mental causation on the grounds that if the mind is a physical entity, then causation already occurs at a physical level, and therefore positing causation at the level of the mental is otiose. This was part and parcel of a general reductionist program [24] which had its own critics [25], but the point is a compelling one. If there is already consciousness at one level of organization of a system, why does that not fully account for, and therefore 'crowd out,' apparent consciousness at other levels? This line of thinking would lead us to conclude that because individual persons integrate much more information than the groups of which they are a part, we ought to think that the individual persons, and not the group, are conscious. Any apparent conscious experience that we might be inclined to attribute to the United States is actually fully explained by the conscious experience of its constituent parts. This line of thinking also explains why we attribute consciousness to the whole brain or to the thalamocortical system rather than a subnetwork or constituent set of neurons: the only *real* action is happening only where the most action is happening.

You may take this to be intuitive or not. On its face, we might already have a problem with such an argument, especially in light of its similarity to causal exclusion arguments. After all, we often feel inclined to say, for example, that the baseball broke the window, and not that the individual particles that compose the baseball broke the window. Perhaps this is a disanalogy with Tononi's principle—causal exclusion principles are usually reductionist in flavor and thus push us down to the most physically basic level of explanation, while Tononi's exclusion postulate takes the metric at the heart of his Integrated Information Theory of consciousness (viz., informational integration) and selects the level of organization on that basis. However, you still may not be convinced that it is

necessary to identify consciousness at only one level of organization of some system, especially in light of the fact that Tononi accepts that cameras exhibit some low but nontrivial level of consciousness [5, p. 405]. If a camera is conscious, and we were able to build a conscious robot that depended on cameras for its visual modality, why do the cameras suddenly stop being conscious? Block [26] pushes a similar kind of concern. If a tiny, conscious creature entered your brain and replaced one of your neurons, then suddenly (according to anti-nesting principles such as the exclusion postulate) you would cease to be conscious (or perhaps the creature would cease to be conscious, depending on which of you integrated more information). The idea that the consciousness of some system could flicker in and out merely on the basis of its relation to other conscious entities seems problematic.[6]

## §3. List's argument against (significant) consciousness of collective entities

Though Schwitzgebel doesn't endorse a specific functionalist account of consciousness in that original paper, he claims in subsequent work that Tononi's Integrated Information Theory of consciousness (IIT) provides evidence for collective consciousness (more specifically, the consciousness of the United States) [28]. It is interesting, then, that List [4] argues, on the basis of IIT, that the United States is *not* conscious. The rest of this paper is focused on navigating this dialectic and, more specifically, introducing some considerations in favor of thinking that the truth of IIT might entail the existence of conscious group agents.

Before I go any further, I should make one thing clear. It is not entirely accurate to say that List rules out consciousness for group entities. Rather, his claim is the following: collective entities of similar functional organization to the United States have *insignificant* levels of consciousness. Thus, he does not entirely disagree with Schwitzgebel's overall point. However, his argument, if sound, serves to deflate it tremendously. On List's account, the U.S. may be conscious, but in the same way that a camera or a photodiode is conscious (which is to say, not very). Thus, to preserve the spirit of Schwitzgebel's bold claim, which is that collective entities are potentially on a par with conscious organisms, we will have to reject List's argument.[7]

---

[6] Kammerer [2] attempts to articulate a sophisticated anti-nesting principle that does not run into the same problems, but Schwitzgebel [27] argues that it is basically unsuccessful, and I endorse his argument. Most germane for our purposes, Kammerer's anti-nesting principle does not seem to rule out collective consciousness *in principle*; at best, he can only hope to show that collective entities that are similar in organization to the United States are not conscious. Other everyday collective entities, on the contrary, are left on the table.

[7] This also seems important for rejecting Baddorf's argument. If groups can be conscious, but only insignificantly so, then they can be minded, but only insignificantly so, and therefore may not be apt candidates for moral responsibility in the form of accountability.

Why does List take IIT to tell against the possibility of significant collective consciousness? In truth, it is because he expresses IIT in finer detail than Schwitzgebel does. Schwitzgebel uses a rather informal notion of informational integration, while Tononi employs an information-theoretic sense of integration in the development of his theory. Tononi uses the symbol Φ to represent the measure of informational integration he has in mind, so I will follow him stylistically.

In basic terms, Φ is a measure of how much information is produced by a system as a whole compared to how much information is produced by the individual components of the system independently. Information is understood "in the classic sense of reducing uncertainty or ignorance" [5, p. 406]. The more information one has, the more possibilities one can rule out. Information, formalized as 'relative entropy,' "is a difference between probability distributions [of potential states of the world before and after some state change in the system]: if the distributions are identical, relative entropy is zero; the more different they are, the higher the relative entropy [information]" [5, *ibid.*]. A system integrates information to the degree that the probability distribution of the system as a whole differs from the probability of the system modeled as independent parts. In other words, the more that the different components of the system interdepend in the production of information, the more the system is functionally integrated (and thus is conscious). Finally, we can take the system as a whole and decompose it in such a way that leaves the least information unaccounted for. If we can decompose the system into constituent parts such that no information is lost, then the system has no informational integration ($\Phi = 0$), and thus, on Tononi's account, has no consciousness taken as a system in its entirety.

The question, then, is whether collective entities are functionally organized such that their constituent parts are highly interdependent. From the opposite perspective, if we can account for a significant proportion of the information generated by a collective entity by considering its components in isolation, the more it looks as though the collective entity is not significantly conscious. This decomposition is what List terms a 'low-entropy partition,' and he believes that such a partition is likely to obtain for most, and potentially all, group agents.

There are three reasons that List thinks a low-entropy partition of group agents is possible. The first is that "Many group agents, such as corporations, states, or other large organizations, are decomposable into functionally relatively self-contained units, which are each internally more interdependent than they are dependent on other units" [4, p. 312]. This seems true enough—all of these involve bureaucracies divided into multiple sub-units that primarily interact with themselves, not with other sub-units. Second, "Even if we identified the 'cortex' analogue of a group agent, say in the

form of its board of directors or its governing assembly or some other central decision-making body, this 'steering unit' within the collective would still retain much of its functioning even if we hypothetically replaced some part of it with a random process" [4, *ibid.*]. I suppose the idea here is that, if there are nine directors sitting on a board, we could swap one out with a machine that votes randomly, and the board would still be able to get along fairly well. This is at least true if the board's actions were done on a majoritarian basis. Finally, "Much of the information processing in a group agent can be attributed to information processing by individuals" [4, *ibid.*]. This seems similar in spirit to the first reason, which is just the general thought that there is much more going on within individual components of a group agent than between components.

In sum, the point is that collective entities are much less conscious than their constitutive components; in fact, they are not very conscious at all. In principle, it might be possible to calculate the Φ of some collective entity so that its level of consciousness can be compared to that of other entities that are less conscious than we. Perhaps the United States is less like a rabbit and more like a worm, or a camera, or a big, fancy photodiode. Moreover, if one endorses the exclusion postulate, as Tononi does both individually and in collaboration [3], then it is *only* the level of organization of a given system with the highest Φ that is conscious. Even without this exclusivity, though, List takes himself to good give reason to be pessimistic about the phenomenal experience of collectivities.

## §4. A defense of collective consciousness

In this section, I will attempt to rebut List's reasons for refuting significant collective consciousness. Before I do so, I have one more thing to say about the exclusion postulate. Tononi and Koch argue for it on the basis of exclusion as an 'axiom of experience,' which means they don't argue for it. They claim that it is axiomatic of subjective experience that "Consciousness is definite, in content and spatio-temporal grain: each experience has the set of phenomenal distinctions it has, neither less (a subset) nor more (a superset), and it flows at the speed it flows, neither faster nor slower" [3, p. 6]. I understand this as the claims that: a given experience is irreducible to its parts, such that it cannot be experienced as less than it is; and a given experience is no more than its parts, such that there is no more to the experience than the subject does, in fact, experience.[8] They take this to entail that there can only be *exactly one* causal structure of a given system that gives rise to consciousness. Hence, the argument for the exclusion postulate really just is a causal exclusion argument.

[8] As we'll see in §5, this recalls recent work on the unity of consciousness. For an extended discussion, see [29].

I don't object to their premises, but here is a reason to think that they don't license their conclusion. A given experience may be definite in the way they describe, but it strikes me as obviously true that a scene can be experienced by multiple subjects.[9] For each of those subjects, there is a definite (not more-than, not less-than) experience of the scene. What is strange to us about collective entities is how they could possibly have an experience that is distinct from the experiences of their constituent members. This is reflected in Tononi and Koch's claim that the experience of the individual causally excludes the experience of the collective. The same experiences are had (and equivalently, the same information is integrated) at the collective level as at the individual level, given that the perspective of the collective strictly depends on the perspectives of its constituent parts. Yet, all the work of integrating that information is done at the individual level. However, it's not the case that the collective has the same perspective or same experience as any individual member. Its perspective is a singular composite of the perspectives of all of its members, and the method of composition need not be simply additive or aggregative. There's no reason to think that the experience of the group agent, as a subject in itself, is not as definite as the individual experiences of its members unless you reject out of hand the notion that the multiple perspectives of the individual can be composed into one.[10] The question circles back to whether there is different information (or a different method of integrating it) available at the collective level than there is at the individual one.

Let's quickly go back to thinking about awareness, which we admit of groups if we admit that they can be agents. If group agents are aware of things, then they are aware of more things than any individual member. They are aware of all the things of which their members are aware, but also more, as long as they are functionally organized in the right way. As a toy example, we might think that only group agents are aware of the statistical mode of the ages of the group's members, even though each member only knows their own age, and thus no individual member is aware of the same thing as the group. The point is simply that the group agent may well be a subject in itself—to presuppose that the group is not a subject is to beg the question—and this is evidenced by the fact that the group seems to have a perspective that is distinct from those of any of its members, even if it depends on

[9] See [30] for some evidence to this effect. The researchers found that, when multiple people watched the same movie and then were asked to report details about it while in an fMRI scanner, patterns of neural activation in higher-order areas were more similar between people recalling the same event than between recall and perception. The authors take this to indicate that there was at least very similar spatial organization of the memories of those experiences, which means that the experience of those memories is very similar across the different participants, more similar than the firsthand experience and the experience of remembering for any subject.

[10] Sam Coleman [31] thinks that it is impossible for subjects to combine in the way I'm suggesting they do. I'll return to this concern and the nature of the combination in §5.

them. If this is the case, then the group's experience is definite in the same way that the individual's experience is definite. No exclusion involved!

Now to List's non-exclusion-based reasons. In general, I want to argue that List's analysis gives us a good idea of which sorts of groups are not significantly conscious, but that this can direct us to group organizations that are likely to be significantly conscious.

The first reason List gives is that many group agents are decomposable into relatively self-contained units. This is true, but it surely depends on the particular group. Groups are often formed around some sort of collective activity—for corporations, it's business; for states, governance; for romantic partnerships, love. Some sorts of activities are easily done by relatively independent sub-units. The business of large corporations is one example—it doesn't particularly matter to PR what HR is doing. On the other hand, there are some activities that require interdependence between participants. Sports are like this. It is true that a player does not need to depend on their teammates to, say, dribble the ball, but they do need to depend on their teammates to complete a pass, and, moreover, there likely needs to be interdependence within a team *taken as a whole* to move the ball up the field.

This points us in the direction of an important question. On Tononi's theory, consciousness is a matter of integrating information. But what information is at issue here? What possible states of the world are relevant? When we are discussing conscious experience—better said, discrete conscious experiences—the relevant information is that which regards the contents of consciousness. When I am consciously experiencing a red ball, it is a matter of integrating information within the thalamocortical structures of my brain *about that object*, not about anything else. Thus, it seems entirely possible that there will be some tasks in which, and some objects for which, a group produces more information as a result of the functional organization of its members than its individual members produce individually.

Take the example of an army, which appears extremely decomposable on its face due to its strict hierarchical organization. When platoons enter a battle, individual soldiers collect information that is both relevant and not to their individual task. But they are also in constant communication between themselves so that the platoon as a whole knows the location of the enemy, stays abreast of its supplies, and so on. They are acting as a unit to reduce uncertainty about states of the world that are relevant to platoon-level tasks. As we get further up the food chain, individual soldiers may be ignorant of how their individual efforts contribute to group-level tasks, but they do make substantial

contributions—and likewise for the experience that arises from those efforts. It is therefore incorrect to conflate the possible experiences of a collective with that of individual components thereof.

At the level of everyday work, it is true that HR does not care about what PR is doing. But at the level of the action that we ascribe to the group agent, the information relevant to that action is being integrated across the various sub-units—HR, PR, accounting, manufacturing, sales, and so on. This is analogous to the functional organization of the brain. Certain neural circuits in the occipital lobe can handle line detection in isolation. They might function perfectly fine on this task with just one source of information coming in and one source of information going out to other circuits. However, the information relevant to the brain-level task, viz., experiencing some object as a unified content of consciousness, is found from integrating the information across all of these different circuits, hierarchically ordered such that they are eventually all put together *as performance on some brain-level task*. We should expect the same with some collective entities, where the overall task can only be accomplished by significant integration across the different components.

One might accept this, but moving on to List's second reason, think that the sub-components of the 'steering unit' that integrates the information from across the collective is relatively inessential. That is, a board of directors could have one of its members replaced by a machine that votes on issues at random, and the company would still get along fine. Here's the thought as I understand it: if groups are conscious, it is really only the head decision-makers, playing a role analogous to the cortex, that produce consciousness. But even within this smaller group, we could replace certain members with a random process, and the group would still function in relatively the same way. This indicates that the state of any of the individual members of the brain trust is not dependent on the others, which means that there is not that much informational integration, and thus the group is not conscious.

This again fails to capture all sorts of groups. Many groups, for example, require joint attention to function properly.[11] A doubles pair in tennis must both attend to the ball, and must be aware that the other is equally attending to the ball, in order to perform their task. If one tennis player were replaced with a random process, the behavior of the other player would necessarily change significantly—if only because they would have to cover the entire court, rather than just their half. In truth, it is not entirely clear to me that (as List claims) a board of directors would be able to function in basically the same fashion if one of its members were replaced by a random process. Since List says

[11] See [32] for a discussion of joint attention that seems especially pertinent to the points I make here.

nothing more about why he thinks this is true, I see no reason to think that the tennis players of my example are not a counterexample to his claim.

Finally, List claims that "computational contribution made by cross-member connections, while non-negligible, is still moderate compared to the computational contribution made by individual cognitive processes" [4, p. 312]. Again, I think that this may not be true, especially considering task-relevant information. In a phenomenon known as transactive memory [33, 34], long-term partners often 'offload' some knowledge to the other partner, so that a 'collective memory' or 'collective knowledge store' develops between the two of them. Considering this kind of situation, it seems appropriate to attribute significant information processing to the interaction between the two partners. In fact, this is the case for all kinds of collaborative groups. In a brainstorming session, it is often the associations between ideas contributed by different members of the group that ultimately count toward the group's performance on the relevant task. Individual members are thinking, to be sure, but they are thinking in tandem, thinking across the boundaries of their own minds. Much of the information that is produced is in the communication of ideas between individuals, not in the individuals' cognitive processes taken in isolation.

Each of these responses is available because of the specific type of agency that we're dealing with when we consider group agents. Unlike biological agents, whose conscious experience is the consequence of evolutionary and ontogenetic development, the functional organization that would secure corporate agents their conscious experience is in large part intentionally established by its constituent members. Humans form groups in order to pursue ends that are out of reach for individual subjects. We must be disciplined in our thought here, because the notion of 'collective consciousness' naturally pushes the imagination toward *Geist*-like group agents that float free of the individual agents that compose them and enjoy the same sort of consciousness that we do. That's not what we're dealing with. If conscious experience is a matter of information, as IIT posits, then the information that is relevant to artificially-created corporate agents will be different than the information that is relevant to our own experience. It will be the information relevant to the tasks and pursuits that are the reason for their existence in the first place.[12] This means that their experience is likely to be very different in

[12] This isn't to say that the tasks to which groups are put are straightforwardly intended by their members at the moment of their creation. Social groups are sometimes formed unknowingly, for ends that are not entirely transparent to all of their members. Those ends might also change over time as members' needs change. But the point here is to say that individuals operate in groups (in the "we-mode", as Tuomela [35] phrases it) only in the context of certain tasks, and that the experience of groups will be tied to their performance of those tasks.

feel to ours—perhaps different enough that we cannot get a firm grip on what it is like to be them—but this is not a reason to think that it does not have a qualitative aspect.

## §5. The possibility of a collective perspective

Taken together, the examples and arguments I have given cast doubt on List's claim, as well as Tononi and Koch's claim, that no collective is likely to be (significantly) conscious. This does not mean, however, that collective consciousness is identical to individual human consciousness. The examples of alien consciousness that Schwitzgebel uses to motivate the claim that the U.S. is conscious are just that—alien. This should not come as a surprise. We can attribute experience to them without having a great idea of what that would be like. The foregoing analysis can help us to gesture in that direction. In closing, I will attempt to rebut an objection that hangs over this project—an objection to the notion that such a thing as a collective perspective from which an entity experiences the world is even possible.

IIT sits in an uneasy relationship with panpsychism about consciousness. It's common for authors to suggest that IIT "entails" or "implies" some version of panpsychism [36]. Tononi & Koch acknowledge the resonance between the two ideas while rejecting a clear link: "in line with the central intuitions of panpsychism, IIT treats consciousness as an intrinsic, fundamental property of reality. IIT also implies that consciousness is graded, that it is likely widespread among animals, and that it can be found in small amounts even in certain simple systems. Unlike panpsychism, however, IIT clearly implies that not everything is conscious" [3, p. 11]. The lack of implication is in large part due to the exclusion postulate. However, making use of the resources of IIT in defending of group consciousness, while simultaneously denying the exclusion postulate, leads to the same problem panpsychists face: the combination problem. For panpsychists, this is the problem of composing our own subjectivity out of the intrinsic subjectivity of fundamental material entities; for my purposes, this is the problem of composing the subjectivity of a group out of individual, human-level subjectivity.[13]

Coleman [31] advances an argument that, if sound, would strictly rule out the possibility of such combination. He claims that 'points of view' or 'perspectives' of multiple subjects cannot combine into a single perspective. This is largely due to his model of combination, which he bases on the formation of a water molecule out of hydrogen and oxygen atoms. There are two defining characteristics of combination, on his view: "the components continue to exist in the whole, but are

[13] Thanks to an anonymous reviewer for drawing my attention to this issue.

intrinsically altered by combining with one another," and the whole "possesses novel systemic powers" [31, pp. 30-31]. If we understand combination in this way, we get a dilemma. If we posit one source of subjectivity after combination, then both components don't continue to exist in the whole, since we started with two sources of subjectivity. But if we maintain that both perspectives persist, leaving a group-level subject with multiple perspectives, then we have failed to constitute a higher-level subject.

Right off the bat, we may be worried about applying this model of combination to group agents, since we're unlikely to think that the conscious experience of an individual is "intrinsically altered" by participating in a group that has its own stream of experience. That experience may be altered *as it is incorporated into the group's experience*, but it need not be deformed in itself in virtue of participating in combination. Or, to put it another way, the individual's perspective need not be altered at the individual level, even if its contribution to the group-level experience involves alteration.

Coleman further motivates the claim of metaphysical impossibility with a thought experiment about two subjects, each of whom are experiencing a unitary field of color, one red, one blue. He presses that we should think of these subjects' experience as "red-to-the-exclusion-of-(blue-and)-all-else" and likewise for blue [31, p. 33]. Yet, a subject cannot combine these two experiences if it is intrinsic to each experience that it does not include anything else (similar to Tononi's point about the definiteness of experience). He notes that we might imagine the higher-level subject experiencing a kind of 'stereoscopic' vision of "two visual fields, one filled with blue, the other with red. But that is one point of view experiencing blue and red conjunctively, nothing like the original two points of view," and hence we've once again failed to really *combine* experiences into a single subjectivity [31, *ibid.*]. My response here will benefit from further support.

Roelofs suggests that Coleman's argument is an instance of 'the argument from phenomenal interdependence,' which they gloss thus: "given phenomenal interdependence, experience sharing is impossible, because if a single experience were part of two different total phenomenal fields, its interdependence with the other members of those fields would require it to have two different phenomenal characters" [37, p. 3210]. They give a fairly simple reply to Coleman's thought experiment, which is of a piece with my preliminary response. First, they deny the 'exclusivity principle,' which states that all experiences which belong to a particular subject can *only* belong to that subject. This is the same move I made above when I suggested that the group's perspective is a composite of its members' perspectives—their experiences are shared in some way by the group. They continue:

> the defender of experience sharing can accept that no composite could have a unified experience of "red-to-the-exclusion-of-(blue-and)-all-else [and] blue-to-the-exclusion-of-(red-and)-all-else." Both experiences would be affected by the other so as to lose their 'to the exclusion of' character. But this can be true also of the parts of the imagined composite: one can experience red, and the other blue, without either experience having the character of 'excluding' the other, precisely because those experiences are unified and thus adjust each other's phenomenal character. Perhaps each does, when isolated, experience their redness or blueness as excluding all else, but once they are connected into a unified whole, which experiences the red as unified with blueness, and vice versa, so do they [37, p. 3213].

This is basically to reiterate the point I made above, namely, that experiences can be altered in combination without being altered as they exist in isolation from one another. Even if you don't find this response compelling in the panpsychist context, I think it should be more compelling in the context of group agents, because whereas we might think of microsubjects who retain an independent existence strange, we don't have the same prejudice about human subjects who participate in groups. Plus, in line with my comments about the functionality of collective consciousness at the end of §4, we should be asking ourselves questions like the following: *Why would these two subjects come together to form a group subject that would share in their experiences in the first place? For what purpose do their experiences need to be integrated at some higher order?* If it is important that the group subject integrate information about those two color experiences, then it will compose them in such a way that they lose their 'to the exclusion of' character at the group level. But if it is not relevant to the purposes for which these two individuals are acting together in a group context, then we do not have to give an explanation for how *that* aspect of their experience combines at all.

In closing, let's try to make this picture more concrete. We want to know something about what collective conscious experience would be like. This will depend on what sort of alterations are made between individual experience and collective experience in the process of the former combining into the latter. I hypothesized that group consciousness is likely to be more task-dependent than human consciousness. If this is right, we can make some further hypotheses about combination.

First, it is likely to be more transient, only arising when individuals are faced with group-level tasks. A corporation, if it is ever conscious, will not be conscious while its employees are not at work.

It will be, if you like, 'asleep' or in 'zombie-mode' outside of working hours. Second, and to the point above, the experience of groups may involve fewer modalities, and probably only those that are important to the task at hand. A group may not 'see' through all of its members' eyes, for example, if integrating visual information is not essential to the task in question. A platoon, on the other hand, might 'see' and 'hear' in one unified stream, because the visual and auditory inputs of all of its members must be communicated effectively in order to perform the task of the group agent. However, it wouldn't be likely to have tactile experiences.

Finally, group consciousness may be impoverished in certain ways, though also potentially more enriched in others. If a group really does 'see' through its members' eyes, then it sees much more that we can ever hope to see as an individual subject. However, the fidelity and richness of experience is likely to be a function of the efficacy of communication between different group members. This, I think, is one of the major intuitive reservations about attributing consciousness to groups. Not only would group consciousness supervene, in some respect, on individual consciousness, but the connections between the different conscious sub-processes of the plural subject would likely be noisier and slower than the connections between unconscious sub-processes of the individual subject.

This is similar in certain ways to the 'stereoscopic vision' response to Coleman's thought experiment, which takes on a new meaning in this light. What happens if the experiences of individual members of a collective are *prima facie* incompatible, as when one member hears a loud noise but another hears perfect silence?[14] The cheap answer is that it will probably depend on the task for the purposes of which the group reaches sufficient informational integration to attain consciousness. Again, think of the platoon. In that case, I think it's plausible that the platoon as a distinct entity hears the loud noise, because its task depends on its ability to orient toward loud noises. In each of these cases, we have seen that the process of combination takes what is needed from individual experience in forming a task-appropriate composite experience. This is only a preliminary approach to detailing what sort of alteration to the individual perspectives this combination would consist in, but hopefully it is sufficient to give some insight into the experience of these strange entities of our own creation that live among us. Or, more reservedly, we at least see why IIT does not rule out such entities having experiences of their own.[15]

[14] Thanks to an anonymous reviewer for pressing this issue.

[15] I am grateful to Caleb Dewey, Eve Isham, Terry Horgan and his Philosophy of Mind working group, Brandon Ashby, and two anonymous reviewers at *Neuroethics* for their comments on this paper at various stages in its production.

References

1. Schwitzgebel, Eric. 2015. If materialism is true, the United States is probably conscious. *Philosophical Studies* 172: 1697–1721. https://doi.org/10.1007/s11098-014-0387-8.

2. Kammerer, François. 2015. How a Materialist Can Deny That the United States is Probably Conscious – Response to Schwitzgebel. *Philosophia* 43: 1047–1057. https://doi.org/10.1007/s11406-015-9653-z.

3. Tononi, Giulio, and Christof Koch. 2015. Consciousness: here, there and everywhere? *Philosophical Transactions of the Royal Society B: Biological Sciences* 370: 20140167–20140167. https://doi.org/10.1098/rstb.2014.0167.

4. List, Christian. 2018. What is it Like to be a Group Agent? *Noûs* 52: 295–319. https://doi.org/10.1111/nous.12162.

5. Tononi, Giulio. 2009. An Integrated Information Theory of Consciousness. In *Encyclopedia of Consciousness*, 403–416. Elsevier. https://doi.org/10.1016/B978-012373873-8.00014-1.

6. Epstein, Brian. 2018. Social Ontology. *Stanford Encyclopedia of Philosophy*.

7. Durkheim, Emile. 1982. *The Rules of Sociological Method*. Edited by Steven Lukes. London: Macmillan Education UK. https://doi.org/10.1007/978-1-349-16939-9.

8. Hegel, Georg Wilhelm Fredrich. 2018. *The Phenomenology of Spirit*. Eds. Terry Pinkard and Michael Baur. Cambridge University Press. https://doi.org/10.1017/9781139050494.

9. Althusser, Louis, and Étienne Balibar. 2015. *Reading capital: the complete edition*. London; New York: Verso.

10. Rupert, Robert D. 2014. Against Group Cognitive States. In *From Individual to Collective Intentionality*, eds. Sara Rachel Chant, Frank Hindriks, and Gerhard Preyer, 97–111. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199936502.003.0005.

11. List, Christian, and Philip Pettit. 2011. *Group Agency*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199591565.001.0001.

12. Björnsson, Gunnar, and Kendy Hess. 2017. Corporate Crocodile Tears? On the Reactive Attitudes of Corporate Agents. *Philosophy and Phenomenological Research* 94: 273–298. https://doi.org/10.1111/phpr.12260.

13. Block, Ned. 1995. On a confusion about a function of consciousness. *Behavioral and Brain Sciences* 18: 227. https://doi.org/10.1017/S0140525X00038188.

14. Kouider, Sid, Vincent de Gardelle, Jérôme Sackur, and Emmanuel Dupoux. 2010. How rich is consciousness? The partial awareness hypothesis. *Trends in Cognitive Sciences* 14: 301–307. https://doi.org/10.1016/j.tics.2010.04.006.

15. Brooks, D.H.M. 1986. Group minds. *Australasian Journal of Philosophy* 64: 456–470. https://doi.org/10.1080/00048408612342641.

16. Huebner, Bryce. 2013. *Macrocognition: A Theory of Distributed Minds and Collective Intentionality*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199926275.001.0001.

17. Pettit, Philip. 2018. Consciousness Incorporated. *Journal of Social Philosophy* 49: 12–37. https://doi.org/10.1111/josp.12219.

18. Chalmers, David J. 1996. *The conscious mind: in search of a fundamental theory*. New York: Oxford University Press.

19. Baddorf, Matthew. 2017. Phenomenal consciousness, collective mentality, and collective moral responsibility. *Philosophical Studies* 174: 2769–2786. https://doi.org/10.1007/s11098-016-0809-x.

20. Putnam, Hilary. 1965. Psychological predicates. In *Art, mind, and religion*, ed. W. H. Capitan and D. D. Merrill. Liverpool: University of Pittsburgh Press.

21. Tononi, Giulio. 2012. The integrated information theory of consciousness: An updated account. *Archives Italiennes de Biologie* 150: 290–326.

22. Kim, Jaegwon. 1993. *Supervenience and mind: selected philosophical essays*. Cambridge Studies in Philosophy. New York, NY, USA: Cambridge University Press.

23. Kim, Jaegwon. 1998. *Mind in a Physical World: An Essay on the Mind-Body Problem and Mental Causation*. The MIT Press. https://doi.org/10.7551/mitpress/4629.001.0001.

24. Kim, Jaegwon. 1992. Multiple Realization and the Metaphysics of Reduction. *Philosophy and Phenomenological Research* 52: 1. https://doi.org/10.2307/2107741.

25. Fodor, Jerry. 2008. Special Sciences: Still Autonomous after All these Years. *Noûs* 31: 149–163. https://doi.org/10.1111/0029-4624.31.s11.7.

26. Block, Ned. 2007. Troubles with Functionalism. In *Consciousness, Function, and Representation*. The MIT Press. https://doi.org/10.7551/mitpress/2111.003.0006.

27. Schwitzgebel, Eric. 2016. Is the United States Phenomenally Conscious? Reply to Kammerer. *Philosophia* 44: 877–883. https://doi.org/10.1007/s11406-016-9725-8.

28. Schwitzgebel, Eric. 2012. Why Tononi should think that the United States is conscious. *The Splintered Mind*. https://schwitzsplinters.blogspot.com/2012/03/why-tononi-should-think-that-united.html

29. Bayne, Tim. 2010. *The Unity of Consciousness*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199215386.001.0001.

30. Chen, Janice, Yuan Chang Leong, Christopher J Honey, Chung H Yong, Kenneth A Norman, and Uri Hasson. 2017. Shared memories reveal shared structure in neural activity across individuals. *Nature Neuroscience* 20: 115–125. https://doi.org/10.1038/nn.4450.

31. Coleman, Sam. 2014. The Real Combination Problem: Panpsychism, Micro-Subjects, and Emergence. *Erkenntnis* 79: 19–44. https://doi.org/10.1007/s10670-013-9431-x.

32. Siposova, Barbora, and Malinda Carpenter. 2019. A new look at joint attention and common knowledge. *Cognition* 189: 260–274. https://doi.org/10.1016/j.cognition.2019.03.019.

33. Wegner, Daniel M. 1987. Transactive Memory: A Contemporary Analysis of the Group Mind. In *Theories of Group Behavior*, eds. Brian Mullen and George R. Goethals, 185–208. New York, NY: Springer New York. https://doi.org/10.1007/978-1-4612-4634-3_9.

34. Wegner, Daniel M., Ralph Erber, and Paula Raymond. 1991. Transactive memory in close relationships. *Journal of Personality and Social Psychology* 61: 923–929. https://doi.org/10.1037/0022-3514.61.6.923.

35. Tuomela, Raimo. 2007. *The Philosophy of Sociality*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195313390.001.0001.

36. Mørch, Hedda Hassel. 2019. Is the Integrated Information Theory of Consciousness Compatible with Russellian Panpsychism? *Erkenntnis* 84: 1065–1085. https://doi.org/10.1007/s10670-018-9995-6.

37. Roelofs, Luke. 2016. The unity of consciousness, within subjects and between subjects. *Philosophical Studies* 173: 3199–3221. https://doi.org/10.1007/s11098-016-0658-7.