What is it like to be a group agent?

Christian List

June-July 2015[1]

The existence of group agents is relatively widely accepted. Examples are corporations, courts, NGOs, and even entire states. But should we also accept that there is such a thing as group consciousness? In this paper, I give an overview of some of the key issues in this debate and sketch a tentative argument for the view that group agents lack phenomenal consciousness, contrary to a recent suggestion by Schwitzgebel (2015). In developing my argument, I draw on integrated information theory, a much-discussed theory of consciousness. I conclude by pointing out an implication of my argument for the normative status of group agents.

1. Introduction

It is, by now, relatively widely accepted that suitably organized collectives can be intentional agents in their own right, over and above their individual members (see, e.g., French 1984; Rovane 1997; Pettit 2001, ch. 5, 2003; List and Pettit 2006, 2011; Tollefsen 2002, 2015; Tuomela 2013).[2] Examples of group agents include commercial corporations, collegial courts, non-governmental organizations, even states in their entirety. Like an individual human being, a group agent has purposes and intentions and pursues these through its actions. In doing so, it can be as rational as an individual rational agent, at least when we understand rationality in the way decision theorists do. For example, a firm's behaviour in the market place can be well understood by modeling it as a rational utility maximizer, and corporations often fit the model of *homo economicus* better – and to a scarier extent – than most individual human beings do.

If we accept that there are group agents, should we also accept that there is such a thing as group *consciousness*? This is the issue I want to explore in this paper. Is there anything it is like to be a group agent? Thomas Nagel famously asked: "What is it like to be a bat?" Nagel's point was that the defining mark of a conscious organism is that "there is something that it is like to *be* that organism" (Nagel 1974, p. 436). There is something it is like to be a human being; you are experiencing it right now. Presumably, there is also something it is like to be a chimpanzee. Similarly, other mammals, such as cats, dogs, and bats, are plausibly conscious: they experience the world, have perceptions, feel pleasure and pain, even if we cannot appreciate what it is like to be in their position. By contrast, rocks, tables, and chairs lack consciousness. There is nothing it is like to be such an entity. So, are group agents more like cats, dogs, and bats in this respect, or more like rocks, tables, and chairs?

---

[1] This paper has been prepared for presentation at several conferences in the summer of 2015. It can be cited as a working paper. For comments and discussion, I am grateful to the audiences at Social Complexes 2, held at the University of Gothenburg, June 2015, and at the 2015 San Raffaele Spring School in Philosophy, held in Milan, June 2015. I also wish to record my intellectual debt to Philip Pettit, with whom I have collaborated on the topic of group agency over many years. For helpful conversations and/or email exchanges, I would further like to thank David Chalmers, Wlodek Rabinowicz, Eric Schwitzgebel, Giulio Tononi, and Laura Valentini.

[2] The thesis that there can be group agents is consistent with what Gilbert (1989) calls *non-singuralism*.

One plausible view, held by a number of philosophers, is that collectives lack consciousness, even when they qualify as agents. In particular, the presence of *intentional* states in a group, such as purposes, goals, and beliefs, is compatible with the absence of *phenomenal* states. Tollefsen (2015, p. 63), for instance, notes that "[i]f we start with a notion of agency that requires … consciousness, we have ruled out groups from the start", but argues that we can defend group agency because "phenomenal consciousness is not required for agency" (p. 52). Similarly, Tuomela (2013, p. 52) says: "A functional group agent has only derived, extrinsic intentionality and, as bodiless, lacks the phenomenal features of normal individual agents". And Theiner (2014) argues that, while groups are capable of cognition, they need not be conscious. He writes: "But what is it like to be a group? Can groups experience the collective equivalent of a headache? If we apply the 'headache criterion' for the existence of minds …, it seems implausible that groups can have mind" (p. 309).[3] The view that groups can have intentions but no consciousness seems consistent with folk-psychological intuitions. As Knobe and Prinz (2007) report, participants in an empirical study were much less willing to attribute conscious mental states to groups than they were willing to attribute intentions without conscious feelings to them.[4]

An alternative view is that we should not rule out the idea of group consciousness so readily. For example, Huebner (2014, p. 120) writes: "it is hard to imagine that collectivities can be conscious; but it is just as hard to imagine that a mass of neurons, skin, blood, bones, and chemicals can be phenomenally conscious … The mere fact that it is difficult to imagine collective consciousness does not establish that absent qualia intuitions have [the force they are sometimes thought to have]."[5] And Schwitzgebel (2015) provocatively argues that "[i]f materialism is true, the United States is probably conscious" (p. 1697), offering perhaps the most elaborate defence of group consciousness in the literature. His central point is that the United States is a sufficiently integrated system such that if we accept materialist criteria for consciousness that are "liberal enough to include both small mammals and highly intelligent aliens, then the United States probably does meet those criteria" (p. 1717).

My aim in this paper is to do three things. First, I want to give an overview of some of the key issues in this debate. Second, I want to sketch a tentative and conditional argument for the view that group agents do indeed lack phenomenal consciousness, contrary to Schwitzgebel's intriguing suggestion. Third, I want to draw attention to one important implication of this view, which concerns the normative status of group agents. My argument is conditional because it depends on a still controversial theory of consciousness: integrated information theory (for an overview, see Tononi and Koch 2015). Although the present paper is largely a review of existing ideas from the literature, I hope that my way of putting them together will be a useful contribution to the debate.

The paper is structured as follows. In Section 2, I briefly introduce the phenomenon of group agency and distinguish it from the related phenomenon of joint agency. In

[3] He cites Harnad (2005), who argues against the possibility of genuine collectively distributive cognition, in support of the "headache test".

[4] However, a study by Huebner, Bruno, and Sarkissian (2010) suggests that "the intuition that there is nothing that it's like to be a collectivity is, to some extent, culturally specific rather than universally held".

[5] In a recent manuscript, Björnsson and Hess (2015) also note that "corporate agents might instantiate various functional properties often associated with phenomenal consciousness".

Section 3, I introduce the notion of consciousness and explain the distinction between consciousness as awareness and consciousness as phenomenal experience (drawing on Chalmers 1995). This will, in turn, allow me to sharpen my question: I will point out that we can unproblematically accept that group agents can have consciousness as awareness, while it is much less clear whether they can also have consciousness as phenomenal experience. In Section 4, I turn to the relationship between functional and phenomenal states of an agent and introduce the notion of a psycho-physical bridge principle, which connects the two. The answer to the question of whether group agents can be phenomenally conscious then depends on which bridge principle is correct. In Section 5, I run through some illustrative such principles and discuss their implications for group consciousness. In Section 6, I focus on one such principle, based on integrated information theory (following Tononi and Koch 2015). Its upshot is that group agents may well lack consciousness, despite being agents in a functional sense. In Section 7, I conclude by discussing an implication of this view, namely that it vindicates a normative asymmetry between group agents and individuals, which was asserted, but not fully defended, in List and Pettit (2011, ch. 8).

2. Group agency

What is a group agent? While we have a decent understanding of what an individual agent is, it may be tempting to define a group agent in some *sui generis* way, for instance by specifying how several individuals must act together in order to form a group agent. However, a more parsimonious approach is to begin with a general definition of an agent, and then to apply it to the case of groups (this is the approach in List and Pettit 2011). A group agent, on that approach, is a collective that qualifies as an agent. The practical question of how a collective can achieve this status – organizationally, institutionally, behaviourally – is separate from the definitional question of what a group agent is.

So, what is an agent? It is helpful to use a functionalist definition, based on the traditional belief-desire model (as in List and Pettit 2011). Functionalism about agency, in very rough terms, is the view that what makes a system an agent – and what generates its intentional states – is nothing heavily metaphysically loaded, but simply the way the system functions, internally and externally, including in relation to its environment. On such a definition, an *agent* is a system that has

- representational states or "beliefs", whose functional role is to depict certain features of the environment as the system "takes them to be";
- motivational states or "desires", whose functional role is to depict certain features of the environment as the system "would like them to be"; and
- a capacity to intervene in the environment on the basis of these states, i.e., to "act" in pursuit of its "desires" in line with its "beliefs".

Human beings, chimpanzees, cats, dogs, and bats all qualify as agents under this definition. Similarly, robots can unproblematically qualify too. Indeed, the conditions are quite minimal, and the systems satisfying them may vary greatly in agential capacities and sophistication. In principle, even a thermostat might qualify as an agent in a very basic sense: it has "beliefs" about the actual temperature, "desires" about the target temperature, and a capacity to "act" by regulating the heating (a well-known observation at least since Dennett's 1987 work on the "intentional stance"). But

nothing much is gained from viewing it this way, since a more straightforward mechanistic explanation is available.[6]

Now, whether a group can meet the conditions for agency depends on how it is organized. A random collective – say the collection of shoppers who happen to be in the supermarket at this moment – lacks the required structure. But a suitably organized collective can in principle meet the conditions. The examples I have mentioned – firms, courts, NGOs, and so on – are all of this kind. Whichever states of an organized collective play the functional roles of beliefs and desires will then qualify as the collective's intentional states: the corporate beliefs and desires.

A simple argument for realism about group agents is a naturalistic indispensability argument. In very rough terms, it can be stated as follows.[7]

Premise 1: Our best social-scientific theories of certain social phenomena – for instance, our best theories of the behaviour of firms in the market place – attribute belief-desire agency of the functionalist kind to (some of) the collectives involved, often by representing them as agents in the decision- and game-theoretic sense.

Intermediate conclusion: According to the naturalistic definition of ontological commitment, those theories are then ontologically committed to group agents.

Premise 2: We should, at least defeasibly, take the ontological commitments of our best scientific theories in any given domain at face value.

Conclusion: We should, at least defeasibly, take our best social-scientific theories' commitment to group agents at face value.

Note that the first premise is an empirical claim about our best social-scientific theories of certain social phenomena. It is undeniable that references to group agents feature prominently in the social sciences, including in some successful and well-confirmed theories, such as the theory of the firm in economics. The intermediate conclusion follows if we apply a thin, Quinean definition of ontological commitment (this definition is further defended in Dietrich and List forthcoming). The second premise expresses the "naturalistic ontological attitude", which has been defended in the philosophy of science and is arguably common among many, though perhaps not all, working scientists (see, e.g., Fine 1984). According to this attitude, our best guide to ontological questions in any given domain lies in our best scientific theories of that domain. The conclusion is a cautious form of realism about group agents.

In order to make sure that this realism about group agents fits with the rest of our scientific worldview, we must still open up the "black box" of any organized collective and ask which internal organizational structures and which mechanisms of aggregation allow the group to function in this way. Much of the analysis in List and Pettit (2011) is devoted to answering this question. For the purposes of this paper, however, I set it aside, and simply note that the argument up to this point has invoked only very thin, functionalist notions, and that our main question about group

[6] For an in-depth application of Dennett-style interpretationism to groups, see Tollefsen (2015).

[7] The present exposition is based on parallel arguments in relation to individual agency in List (2014) and Dietrich and List (forthcoming), though the broad structure is familiar from the philosophical literature.

consciousness has been left open by what I have said. No particular answer is presupposed or implied.

Before turning to this main question, it is important to distinguish the phenomenon of group agency from the related phenomenon of joint agency. We speak of *joint agency* whenever two or more individuals engage in some joint action. For example, they go for a walk together, carry a piano downstairs together, or undertake some common project. Clearly, some shared or joint intentions (or intentions that are suitably collectively directed) need to be present among the participants in order to make joint actions possible. A rich literature analyses this phenomenon (see, e.g., Gilbert 1989; Searle 1995; Bratman 1999, 2014; Tuomela 2007).

What I want to emphasize is that joint agency is not the same as group agency (as argued in detail in Pettit and Schweikard 2006). Two or more individuals who engage in a joint action do not necessarily bring into existence a group agent. We can make sense of joint actions without ascribing a single "centre" of belief, desire, and agency to the group in its entirety. It is sufficient for a joint action that the bearers of the underlying intentions are individuals; no group-level intentional states are needed.

Of course, there may be joint actions within the context of a group agent. The practical mechanisms of forming a group agent may often include shared or joint intentions among the individual members. But joint agency is neither sufficient, nor (from purely logical perspective) even necessary, for group agency.

It should be clear, then, that when I investigate whether there is such a thing as group consciousness I do not refer to the conscious experiences of the individual members of a group, nor to the experiences that might go along with participating in joint action. Gilbert (1989, p. 223), for instance, discusses "feelings of unity" among the members of certain social groups and raises the possibility that a "consciousness of precisely this unity among the members" may be present in some groups. My focus here is on the question of whether the group *as a whole* can have such a thing as consciousness.

3. Consciousness

What is consciousness? We use the term "consciousness" to refer to a number of distinct phenomena. On the one hand, we use it to refer to a set of phenomena related to *awareness*, and on the other hand, we use it to refer to a set of phenomena related to *experience*. We need to begin by disambiguating those two uses of the term; I will do so following the account in Chalmers (1995, 1996).

On the awareness side, we say that someone is "conscious" of a given fact or piece of information to indicate that he or she is *cognitively aware* of that fact or information, meaning roughly that he or she knows it or believes it explicitly and is able to access it cognitively. Similarly, we speak of someone's "conscious state" when we wish to refer to everything that he or she is currently aware of, especially everything that currently falls under the scope of his or her attention. And we say that someone is "conscious" to indicate that he or she is not asleep or comatose, but "awake" and cognitively aware of his or her situation.

Awareness is a functionalist notion. It can be explicated in ordinary, third-personal scientific terms. We can devise functional tests for awareness and attention, for instance by considering a subject's behavioural responses to certain cognitive tasks and by observing his or her interactions with the environment. Neuroscientific data may give us additional clues to the neurophysiological correlates of awareness.

On the experience side, by contrast, we speak of someone's "consciousness" to refer to what he or she subjectively experiences: what it is like to be that agent, from the first-personal point of view. This includes "the felt quality of redness, the experience of dark and light, the quality of depth in a visual field", "the sound of a clarinet, the smell of mothballs", bodily sensations such as pleasures and pains, the subjective quality of emotions, and so on (Chalmers 1995). Philosophers also use the terms "phenomenal experience" or "qualia" to refer these subjectively experienced, first-personal states.

Unlike awareness, phenomenal experience is not a functionalist notion and does not easily lend itself to an ordinary, third-personal scientific analysis. As Nagel (1974, p. 436) already noted:

> "It is not captured by any of the familiar … reductive analyses of the mental, for all of them are logically compatible with its absence. It is not analyzable in terms of any explanatory system of functional states, or intentional states, since these could be ascribed to robots or automata that behaved liked people though they experienced nothing."

Nagel put his finger on something that later led to Chalmers's distinction between the "easy" and "hard" problems of consciousness (1995). The "easy" problems are to explain the structure of awareness and the various phenomena related to it, for instance "the ability to discriminate, categorize, and react to environmental stimuli; the integration of information by a cognitive system; the reportability of mental states; the ability of a system to access its own internal states; the focus of attention; the deliberate control of behavior; the difference between wakefulness and sleep".

What makes the "easy" problems easy is not that it does not take time, patience, hard work, and ingenuity to explain the phenomena in question. Of course, it does, and scientists deserve to win major prizes for relevant discoveries. What makes them easy is that, being essentially functional phenomena, they are amenable to an ordinary scientific analysis, using the tools of a broadly physicalist science.

The "hard" problem, by contrast, is to explain phenomenal experience itself. We need to explain the following: why are we not merely functional systems which, for example, form beliefs about red objects, say for distinguishing ripe from unripe tomatoes, so that we can eat the former but not the latter? Why is there something it feels like to experience the bright red of a perfectly ripe tomato? In short, we need to explain why there is something it is like to be us, why we have phenomenal states at all, as opposed to merely functional states.

As Chalmers notes, "[w]hat makes the hard problem hard and almost unique is that it goes beyond problems about the performance of functions" (1995). While most other phenomena studied in the sciences are of a functionalist kind, a purely functionalist account of an agent cannot explain, even in principle, why certain phenomenal states

accompany the functional ones. There seems to be no *logical* contradiction involved in postulating an agent who is functionally indistinguishable from an ordinary human being, who is even indistinguishable with respect to everything that has to do with awareness, but who lacks any phenomenal experience. Such an agent is called a *zombie*. The point is that the notion of a zombie is logically coherent, even if it turns out that there are no zombies in the actual world. And the very coherence of that notion is enough to illustrate that the phenomenal facts about an agent, if there are any, are not simply subsumed by the functional facts, but go beyond them.

To summarize: there is an important distinction between consciousness as awareness and consciousness as experience. The former, but not the latter, is a functionalist notion.

This allows me to draw a first, preliminary conclusion. It is that group agents can certainly have consciousness as awareness.[8] We can meaningfully talk about which pieces of information a group agent such as the FBI is aware of in an investigation (Goldman 2004); and we can give an ordinary functionalist analysis of what we mean by awareness here. For instance, something on which the group agent holds an explicit belief, and which is accessible and reportable, falls under the umbrella of its awareness. We can also meaningfully talk about which things a group agent attends to or fails to attend to. For instance, an organization that has been struck by some scandal might give its attention to this issue and act so as to become clean; this could involve organizationally endorsing and enacting a new policy or code of conduct. We can even make sense of the idea of perceptual awareness in a group agent. As an information processing system, a group agent has various routes of epistemic access to the world. These are mediated through its individual members and its procedures, just as an individual agent's perception is mediated through its sense organs and cognitive processes. And they may be sensitive to some features of the environment but not to others. Just as we humans are sensitive to sounds at certain frequencies but not to sounds outside that range, so a group agent may be perceptually sensitive to some environmental features but not to others.

It should be evident that all of these phenomena can be analysed in ordinary functionalist terms. And the awareness capacities listed by Chalmers, such as "the ability to discriminate, categorize, and react to environmental stimuli", "the integration of information by a cognitive system", "the focus of attention", and "the deliberate control of behavior", can in principle be found in group agents as much as they can be found in individuals.

Even the notions of "wakefulness and sleep" make sense in the context of a group agent. My university goes on vacation from time to time, which means that all offices and institutional activities are closed, all email servers go into vacation-response mode, and all official business is put on hold until the end of the break. The group agent will "wake up" and respond during that break only in a real emergency, such as a scandal suddenly uncovered by the press. This parallels the way in which a sleeping person or animal may wake up in a threatening situation.

[8] Relatedly, Tollefsen (2015) argues that intentional agency requires what Block (2008) calls "access consciousness" (which falls under the rubric of awareness) but not phenomenal consciousness. The implication is that group agents can indeed have access consciousness.

My claim that group agents can have consciousness as awareness is not just metaphorical. Rather, we can be realists about a group's awareness, using the resources of functionalism about agency. By contrast, it is much less clear whether group agents could also have consciousness as experience. To address this question, I need to say more about the relationship between functional and phenomenal states. I will first discuss this question in general, before turning to the case of group agents.

4. The relationship between functional and phenomenal states

How, then, are an agent's functional states related to its phenomenal states, if there are any? There are broadly three possible ways of responding to this question. The first amounts to a denial that an answer is needed. Instead, the suggestion is that once we have fully explained an agent's functioning, we have explained all there is to be explained. On this picture, phenomenal consciousness – over and above functional awareness – is just a powerful illusion. This is roughly Dennett's view, though I have here stated it in slightly simplified terms (see Dennett 2005). The view, which is sometimes called *eliminative* (or *type-A*) *materialism*, is interesting, but I will set it aside, on the grounds that it does not do justice to the "what is it like" question.

A second answer to our question acknowledges that there is a "gap" between an agent's functional and phenomenal states (Levine 1983), and hence that something needs to be said about the relationship between the two, but asserts that phenomenal states, where they occur, supervene on functional states. On this picture, the apparent gap between functional and phenomenal states lies, at least in part, in the fact that we use very different concepts to describe these two kinds of states, but it is nonetheless true that once all the facts about an agent's functional states are fixed, this also fixes all the facts about its phenomenal states, as a matter of necessity. Phenomenal states, wherever they occur, are thus supervenient on functional states, though not reducible to them. This view is sometimes called *a posteriori* (or *type-B*) *materialism*.

The third answer to our question insists that the gap between an agent's functional and phenomenal states is stronger than acknowledged by type-B materialism. According to this answer, the phenomenal states are not a logically or metaphysically necessary byproduct of the functional states, but accompany them only relative to some law of nature – one that is in place in our world, but that is still contingent: postulating a world without that law is free from contradiction. This last view is Chalmers's, and it can be described as a kind of *naturalistic dualism*. (For a defence, see Chalmers 1996.)

Naturalistic dualism can still uphold a more modest supervenience thesis: the phenomenal states of an agent, where they occur, *nomologically* supervene on its functional states, i.e., they supervene on them relative to the relevant law of nature. This is weaker than *metaphysical* supervenience, insofar as the functional states would not be sufficient to give rise to the phenomenal states in the absence of the relevant law.

However, there is one important thing that type-B materialism and naturalistic dualism have in common: they both take the relationship between functional and phenomenal states to be non-trivial. According to both views, there is an important question to be asked about *when exactly* a system's functional states are such as to

give rise to phenomenal consciousness. What conditions must a system's functional states satisfy in order to render the system phenomenally conscious?

Let me call a specification of those conditions a *psycho-physical bridge principle*. Such a principle might take the form of a biconditional: "the system is phenomenally conscious if and only if it has such-and-such functional properties". Alternatively, it might take a weaker form, specifying only sufficient conditions, or only necessary conditions, for phenomenal consciousness. Or finally, the principle might make phenomenal consciousness a matter of degree, saying something along the lines: "the system is phenomenally conscious to the degree captured by such-and-such quantitative property".

According to both type-B materialism and naturalistic dualism, the quest for a psycho-physical bridge principle is a meaningful exercise. The two views only disagree about the status of such a principle. According to type-B materialism, the true psycho-physical bridge principle – whatever it turns out to be – will be a necessary truth: the connection between functional and phenomenal states could not have been otherwise, not even in a hypothetically different universe. Discovering the correct bridge principle is like discovering necessary truths elsewhere, such as in logic, mathematics, or semantics. According to naturalistic dualism, by contrast, the correct psycho-physical bridge principle is a contingent truth about our world, akin to other laws of nature. Discovering it is like discovering the laws of gravity or the laws of electromagnetism.

The question of whether there is anything it is like to be a group agent then depends on which psycho-physical bridge principle is true. In what follows, I will run through a number of illustrative such principles and discuss their implications for group consciousness. While many of those principles are ultimately unsuccessful – and covered here mainly for pedagogical reasons – others are more promising. The final principle, in particular, has some promise and has recently been much discussed in neuroscience, though it remains controversial and is probably still too crude in its present form.

5. Some illustrative psycho-physical bridge principles and their implications for group consciousness

The principles to be discussed can be grouped into three sets. They differ in what they take to be the functional correlates of phenomenal consciousness. According to the principles in the first set, phenomenal consciousness is tied to intentional agency and/or cognition. According to those in the second, it is tied to biological brain activity. And according to those in the third, it is tied to information processing more generally.

5.1 Intentional agency and/or cognition as a correlate of consciousness

A very simple bridge principle is the following:

Principle 1a: Wherever there is intentional agency and/or cognition, there is phenomenal consciousness.

This is extremely permissive, implying that phenomenal consciousness occurs not only in humans, chimpanzees, and bats, but also in robots, thermostats (provided they qualify as agents or cognitive systems), and indeed group agents. And if we take the set of cognitive systems to be larger than the set of intentional agents – maybe because we take cognition to be less demanding than full-blown agency – then the principle will imply that there are even further bearers of consciousness.

The problem with this principle is that while it may not entail many, if any, false negative verdicts about consciousness, it might well entail a lot of false positives. Are we really confident that simple robots or thermostats are conscious? And if a computer can plausibly be described as a domain-specific cognitive system, is it also phenomenally conscious? What evidence could we have for any of these claims?[9]

Perhaps the following more demanding principle is more plausible:

Principle 1b: Wherever there is intentional agency and/or cognition *above a certain level of complexity*, there is phenomenal consciousness.

If we set a sufficiently high threshold for the required level of complexity, this principle will avoid some of the apparent false positives of the earlier principle, but at the risk of being *ad hoc*. What exactly is the relevant threshold? It cannot be related too closely to the cognitive capacities of humans, such as language, because we would otherwise have to conclude that non-human mammals lack consciousness, which is implausible. Further, the principle does not capture the idea that consciousness comes in degrees, and that there is "more" consciousness in complex agents such as human beings than in simpler ones such as cats or mice. However, the present principle might lead us to conclude that certain group agents have consciousness, provided they inherit the relevant cognitive capacities from their human members.

One version of the cognitive-complexity-based principle for consciousness links consciousness to certain forms of higher-order cognition:

Principle 1c: Wherever there is higher-order cognition, there is phenomenal consciousness.

There are different ways of spelling out the notion of "higher-order cognition", but all variants refer to certain higher-order representations of first-order mental states (for a comprehensive survey and discussion, see Carruthers 2011). A higher-order representation of the relevant sort could be an agent's (or cognitive system's) *perception* of its first-order mental states, for instance its perception that it has certain beliefs or desires. Or it could be a higher-order *belief* about those first-order states, or another kind of *thought* or *self-representation*. Moreover, the condition for consciousness could be either the actual presence of the relevant higher-order representation or merely the disposition to form it in appropriate conditions.

[9] On a more demanding view about intentional agency and/or cognition than the simple functionalist view adopted here, one might be able to shrink the set of false positives. In the limit, if we were to build consciousness into our definition of agency and/or cognition, then the presence of agency or cognition would immediately entail the presence of consciousness. But this would not be illuminating. Furthermore, one might worry that such a more demanding view about agency or cognition is less useful and flexible for scientific purposes than the "thin" functionalist view adopted in this paper.

Independently of how we settle these details, it should be clear that, just as human beings are capable of higher-order cognition, there is no barrier to a group's being organized so as to engage in this sort of cognition too. Indeed, it should be possible to come up with functionalistically impeccable criteria for higher-order group cognition.

However, even if we set aside the question of *why* higher-order cognition should be associated with phenomenal consciousness, a serious objection to such an account of consciousness is that it fails to vindicate the presence of consciousness in non-human animals. As Carruthers (2011) notes:

> "Since there is considerable dispute as to whether even chimpanzees have the kind of sophisticated 'theory of mind' to enable them to entertain thoughts about experiential states as such …, it seems most implausible that many other species of mammal (let alone reptiles, birds, and fish) would qualify as phenomenally conscious, on these accounts. Yet the intuition that such creatures enjoy phenomenally conscious experiences is a powerful and deep-seated one, for many people."

Ideally, a good bridge principle for consciousness should capture all paradigm cases adequately: it should entail, for instance, that there *is* consciousness in humans, chimpanzees, and dogs; and that there is no consciousness in tables, chairs, and combustion engines. And it should account for these cases in a systematic and non-*ad-hoc* way. Furthermore, the principle should be revisionary at most in intuitively marginal cases – and if so, for compelling reasons.[10]

Arguably, the previous three principles do not meet these desiderata, and so we should treat their implications for group consciousness with caution. This leads me to turn to the next set of principles, which focus on brain biology.

5.2 Biological brain activity as a correlate of consciousness

Again, we begin with a very simple principle:

Principle 2a: Wherever there is a living mammalian (or similar) brain, there is phenomenal consciousness.

If it turned out that humans and non-human animals with mammalian or similar brains are the only creatures that are, as a matter of fact, phenomenally conscious, the present principle might even be approximately adequate, more-or-less correctly picking out all actual instances of consciousness. But it would still leave a number of important questions open. Is it not possible that phenomenal consciousness could be present in organisms with a radically different brain biology?[11] And what is it about brains that gives rise to consciousness? Is it something about their biological make-up, or rather something about their functional organization? What about a computer simulation of a biological brain? Would this support phenomenal consciousness too?

---

[10] I would suggest that, in the absence of any countervailing considerations, the present desiderata seem reasonable preconditions that a bridge principle should meet before we can justifiably use it to reach any verdicts about non-standard cases, such as the case of group consciousness.

[11] Schwitzgebel (2015) rightly notes, for example, that we would not wish to rule out consciousness in intelligent aliens simply by stipulation.

By tying consciousness so closely to brain biology, a principle like the present one would exclude the possibility of group consciousness, except perhaps in those thought experiments in which a large population of people enact a computational simulation of a biological brain (e.g., Block 1980). However, even in the case of human consciousness, the present principle is unsatisfactory. It does not tell us anything, for instance, about what distinguishes the living brain in a healthy and fully awake grown-up from the living brain in a comatose patient or even in someone who is asleep. A more refined principle is the following:

Principle 2b: Wherever there is a living mammalian (or similar) brain with suitably synchronized patterns of neural activity, there is phenomenal consciousness.

A version of this principle was famously proposed by Crick and Koch (1990), who argued that certain kinds of synchronized neural firing patterns, especially in the 40 Hz range, are associated with consciousness, though in more recent work they no longer endorse this as a sufficient condition (Crick and Koch 2003). Relatedly, Clark (2009) discusses the possibility that "conscious experience requires cortical operations that involve extremely precise temporal resolutions, such as the synchronous activation of distinct neural populations where the required synchrony demands millisecond precision" (p. 984).

The problem with the principle that links consciousness to synchronized neural activity is at least twofold. First, it is highly specific to the implementation of consciousness in the biological brain. Hence, it is either inapplicable to systems that are very different from the biological brain, ranging from robots to group agents, or, even if deemed applicable, it implies – without much explanation – that those systems cannot have consciousness. Second, even in the case of the biological brain, the principle does not seem to be fully adequate. As Tononi and Koch (2015, p. 10) note, consciousness is "lost during generalized seizures, when neural activity is intense and synchronous". So, synchronous neural activity cannot be the full story.[12]

The limitations of the bridge principles based on biological brain activity lead us to move on to another set of principles, which focus on information processing, independently of the biological hardware instantiating it.

5.3 Information processing as a correlate of consciousness

Chalmers (1995) proposed what he called the "double-aspect view of information". Roughly speaking, this asserts that information, wherever it is encoded, has both a functional and a phenomenal aspect. The functional aspect consists in the causal role that information plays in certain physical processes; the phenomenal aspect consists in the (at least tiny amount of) consciousness it gives rise to. At a first gloss, we then arrive at the following bridge principle:

Principle 3a: Wherever there is information processing, there is phenomenal consciousness.

[12] Similarly, Schwitzgebel (2015) notes: "If consciousness, in general, as a matter of physics or metaphysics, requires massive, swift parallelism, then maybe we can get mammal consciousness without U.S. consciousness" (pp. 1710-1711). But he then goes on to draw attention to the limitations of the neural-synchrony thesis about consciousness.

This principle supports a form of panpsychism. Since there is an abundance of information processing in the universe, there is, then, also an abundance of consciousness. There is certainly information processing going on in systems ranging from humans to thermostats, and so all of these systems will have some consciousness. This will be true, in particular, of group agents. Schwitzgebel (2015) also argues that information processing in a collective system might underpin group consciousness.

Like the earlier principle that tied consciousness to agency or cognition, however, the present principle may also seem too unrestricted, and there are a number of issues it does not satisfactorily address. Is consciousness continuously spread out across all of the informationally rich universe, or is it somehow concentrated at certain *loci*, for instance the *loci* of brains and other cognitive systems? And how can we make sense of the *unity* of the consciousness, which is central to our conscious experience? Is the unity of consciousness simply the sum of little bits of consciousness present in all the different informational processes within the human brain?

This suggests that, even if the idea of relating consciousness to information processing is on the right track, we need to pin it down further. I will now review two prominent bridge principles that attempt to do this, by taking the correlate of consciousness to be, not information processing *simpliciter*, but *integrated* information processing.

One proposal comes from "global workspace theory" and predates Chalmers's double-aspect view of information (Baars 1988):

Principle 3b: Wherever an information processing system involves a "global workspace" that integrates and redistributes information from multiple sources, there is (phenomenal) consciousness.

Baars (2003) elaborates this idea as follows:

> "Global Workspace theory suggests a fleeting memory capacity that enables access between brain functions that are otherwise separate. This makes sense in a brain that is … a massive parallel distributed system of highly specialized processors. In such a system coordination and control may take place by way of a central information exchange, allowing some specialized processors – such as sensory systems in the brain – to distribute information to the system as a whole. This solution works in large-scale computer architectures, which show typical 'limited capacity' behavior when information flows by way of a global workspace."

Global workspace theory captures some salient features of consciousness, including its apparent unity, and suggests a functional role for consciousness. Moreover, it does in principle permit group consciousness. All that a group agent would need to have in order to count as conscious is a "global workspace" that serves to integrate and redistribute information in the right way.[13] This conclusion is in line with the

[13] Schwitzgebel (2015, p. 1712), too, mentions global workspace theory in his argument for group consciousness.

observation that group agents can unproblematically have "access consciousness", which falls into the category of consciousness as awareness.

What remains unclear, however, is whether global workspace theory truly offers conditions for consciousness as experience. The worry is that the theory simply asserts functionalist conditions for access consciousness and then stipulates, without further explanation, that these are also conditions for phenomenal consciousness. That's why I have bracketed the word "phenomenal" in the theory's bridge principle. As Chalmers (1995) notes, "nothing internal to the theory *explains* why the information within the global workspace is experienced. The best the theory can do is to say that the information is experienced because it is *globally accessible*."

This leads me to consider another bridge principle based on the idea of informational integration. It seeks to account for phenomenal consciousness directly, not via an account of access consciousness. The principle comes from a recent theory of consciousness proposed by the neuroscientist Giulio Tononi and his collaborators under the name "integrated information theory" (see, e.g., Tononi and Koch 2015). Unlike global workspace theory, it focuses, not on informational integration at the cognitive-psychological level, where the information that is being integrated is one of which the subject is aware, but on informational integration at the physical level, where information is simply a feature of a physical system, which can be captured by Shannon's classic notion of entropy.

Integrated information theory asserts the following bridge principle:

Principle 3b: Phenomenal consciousness is associated with integrated information processing in a physical system, and the system's level of phenomenal consciousness increases with its level of informational integration.

In what follows, I will say more about this theory, and explore its implications for group consciousness.

6. Integrated information theory

According to integrated information theory, as just noted, phenomenal consciousness comes in degrees, and a system's level of phenomenal consciousness depends on its level of informational integration, understood in physical rather than cognitive-psychological terms. To explain this further, I will first introduce the relevant notion of integrated information; I will then briefly describe some evidence for its significance as a correlate of phenomenal consciousness; and I will finally turn to the case of group consciousness.

6.1 What is integrated information?

Integrated information, as understood here, is a feature of a physical system. To define it, we must first give a formal description of the system. I will follow the exposition in Aaronson (2014), which I find particularly congenial, though Aaronson himself does not endorse the theory.[14]

[14] For other expositions, see Masafumi, Albantakis, and Tononi (2014); Tononi (2015); and Tononi and Koch (2015).

Suppose that the system consists of n components, and its overall state at any given point in time takes the form of an n-tuple $x = (x_1, x_2, \ldots, x_n)$ in some state space S^n. For each i, x_i is the state of the i^{th} component of the system. For simplicity, let us assume that the states in S are binary. A given component could thus be a simple switch, which may be in an "on" or an "off" state, represented by $x_i = 1$ and $x_i = 0$, like a neuron that might be firing or not. The dynamics of the system can be captured by a state-transition function f from S^n into itself, which assigns to each state $x = (x_1, x_2, \ldots, x_n)$ its successor state $y = f(x) = (y_1, y_2, \ldots, y_n)$.

To get an intuitive grasp of this definition, it is helpful to consider a few examples of state-transition functions. A very simple such function is the one that only ever reverses each component's state (i.e., swaps 0 and 1) such that, for each i, we have $y_i = 1$ if $x_i = 0$ and $y_i = 0$ if $x_i = 1$. Another, slightly less trivial example is the function that sets y_i to be x_{i-1} when $i > 1$ and that sets y_1 to be equal to x_n. Under this state-transition function, information "travels" rightwards through the system. Each component's state at time 1 becomes the adjacent component's state at time 2, with the further stipulation that the n^{th} component is connected up with the first.

A broader class of state-transition functions can be obtained by assuming that the system's n components are the nodes of some network, with connections between some but not all nodes, such that the i^{th} component's post-transition state y_i depends on the pre-transition states of its network neighbours, but not on those of its non-neighbours. Formally, y_i depends on all x_js where there is a connection between the i^{th} and j^{th} components.

To define the level of informational integration in the system, we must ask "whether the x_i's can be partitioned into two sets A and B, of roughly comparable size, such that the [state transitions of] the [components] in A don't depend very much on the [components] in B and vice versa" (Aaronson 2014). If no such partition exists, then the system exhibits a high level of informational integration. On the other hand, if there exists such a partition, then the level of informational integration is much lower.

In our simple example of a system whose post-transition state is always the component-wise reversal of its pre-transition state (a swap of 0 and 1), there is no informational integration whatsoever: each component only interacts with itself, not with other components. By contrast, in the other examples of state-transition functions I have given, there is typically some interdependence between the system's components.

We can formalize this way of measuring informational integration as follows. For any partition of the system's n components $\{1, 2, \ldots, n\}$ into two non-empty subsets A and B, we write x_A and x_B for the sub-tuples of the system's state x, restricted to the components in A and in B, respectively.[15] We can then apply the state-transition function f to pairs of the form (x_A, x_B) and consider the resulting pairs of the form $(y_A, y_B) = f(x_A, x_B)$. Now consider the following hypothetical stochastic process.

Suppose the sub-tuple x_A is given by some input state of our system, but the sub-tuple x_B is drawn randomly from a uniform distribution. How much entropy – i.e., disorder in Shannon's information-theoretic sense – will there be in the output y_A under this

[15] Formally, x_A and x_B are elements of S^A and S^B, respectively.

stochastic process? Slightly more precisely, let Pr be a uniform probability distribution over the state space S^n. Let X_A and X_B be the random variables that determine the input states of the components in the sets A and B under this distribution, and let Y_A be the random variable that generates the output states of the components in A. We are then interested in the conditional entropy of Y_A, given a fixed specification of X_A (i.e., $X_A = x_A$) while X_B is random, formally $H(Y_A \mid X_A = x_A)$.[16] More generally, we can compute $H(Y_A \mid X_A)$ as the probability-weighted average of $H(Y_A \mid X_A = x_A)$ for all possible specifications of x_A.[17]

As just defined, $H(Y_A \mid X_A)$ is the conditional entropy in the A-components, given non-random specifications of the A-components but random specifications of the B-components. This quantity can be interpreted as a measure of how much the state-transition of the A-components depends on the B-components. In particular, if the entropy in the A-components is low despite the randomization of the B-components, the A-components do not depend much on the B-components. For instance, in our trivial system based on component-wise state reversals, the conditional entropy $H(Y_A \mid X_A)$ is zero. By contrast, in systems with greater interdependence, it is higher. Aaronson (2014) denotes that quantity $\mathrm{EI}(B \rightarrow A)$. In the same way, we can define $\mathrm{EI}(B \rightarrow A)$.

Intuitively, a partition of the system's components into two sets A and B displays the least amount of informational interdependence if it minimizes the sum $\mathrm{EI}(B \rightarrow A)+\mathrm{EI}(A \rightarrow B)$, normalized by division by the minimum of the sizes of A and B. The value of $\mathrm{EI}(B \rightarrow A)+\mathrm{EI}(A \rightarrow B)$ for the partition that solves this minimization problem can then be taken to be a measure of the system's overall informational integration. Tononi calls this measure Φ. (Note that there exist some other subtly different definitions; the details do not matter for my purposes here.[18])

6.2 Integrated information as a correlate of phenomenal consciousness

What is the evidence for the bridge principle according to which integrated information, as captured by Φ, correlates with phenomenal consciousness? Two pieces of evidence stand out, though it is important to note that integrated information theory remains controversial (see Aaronson 2014).

First, the principle seems to explain why the cerebral cortex produces consciousness, while – for all we know – the cerebellum does not, although the cerebellum has an even larger number of neurons than the cerebral cortex (Tononi and Koch 2015). The explanation, according to integrated information theory, lies in the different functional organization in these two distinct regions of the brain. Even though we can currently at most give estimates of Φ, it can be argued that the cortex generates a much higher level of informational integration than the cerebellum. And so the notion of informational integration seems to be able to account for the difference in consciousness.

---

[16] Formally, we have $H(Y_A \mid X_A = x_a) = \Sigma_{y_A \in S^A} -\Pr(Y_A = y_A \mid X_A = x_A) \log_2 \Pr(Y_A = y_A \mid X_A = x_A)$.

[17] Formally, we have $H(Y_A \mid X_A) = \Sigma_{x_A \in S^A} \Pr(X_A = x_A)\, H(Y_A \mid X_A = x_a)$.

[18] In following Aaronson (2014), I have not yet used what Oizumi, Albantakis, and Tononi (2014) call "IIT 3.0". I believe that, for my present argument, the current, simplified definition of Φ suffices.

Second, the principle seems to be able to account for some empirical evidence concerning wakefulness, sleep, coma, and anaesthesia. Why does an awake person experience consciousness, while a sleeping person does not, except when dreaming? Why is there a difference between deep sleep and the kind of shallow sleep that involves dreams? And how does general anaesthesia remove consciousness? According to integrated information theory, "the loss and recovery of consciousness should be associated with the breakdown and recovery of the brain's capacity for information integration" (Tononi and Koch 2015, p. 9). As Tononi and Koch point out:

> "This prediction has been confirmed using transcranial magnetic stimulation (TMS) in combination with high-density EEG in conditions characterized by loss of consciousness … If a subject is conscious when the cerebral cortex is probed with a pulse of current induced by the TMS coil from outside the skull, the cortex responds with a complex pattern of reverberating activations and deactivations that is both widespread (integrated) and differentiated in time and space (information rich) … By contrast, when consciousness fades, the response of the cortex becomes local (loss of integration) or global but stereotypical (loss of information)" (ibid.).

Generally, informational integration should be correlated with what Tononi and Koch call the "perturbational complexity index", which measures certain patterns of neural reverberations in response to stimuli. They point out that this measure "decreases distinctly in all the different conditions of loss of consciousness and, critical for a clinically useful device, is high instead in each conscious healthy subject or neurological patient tested so far" (ibid.).

6.3 What are the implications of integrated information theory for group consciousness?

Integrated information theory is one of the first evidence-based theories of phenomenal consciousness that can, at least in principle, say something non-*ad-hoc* about systems that are very different from us. This is because informational integration is defined in a way that is not tied to any particular kind of hardware, such as the biological brain. Rather, we can assess the level of informational integration even in radically different systems, from electronic to collective. Many systems have non-zero Φ and therefore a tiny bit of consciousness according to integrated information theory.

Calculating the precise value of Φ for any given system is difficult. In fact, Aaronson (2014) conjectures that it is a computationally hard problem (which suggests in particular that it is not generally feasible in polynomial time). Furthermore, to perform this calculation, we would need to know the system's exact "wiring diagramme", which is not easy to specify for complex systems such as the brain. However, at least in cases where there exists such a wiring diagramme, Φ is a well-defined quantity. Moreover, since Φ is defined in physical terms, its definition does not depend on any ascription of cognitive or psychological states to the system.

Generally, heuristic considerations may allow us to infer what kinds of systems tend to have high values of Φ and what kinds of systems tend to have lower values. As

Tononi and Koch (2015) note, for example, systems with rich internal feedback loops tend to have higher values of Φ, while "feed-forward" systems have lower or even zero values. In a pure feed-forward system, "one layer feeds the next one without any recurrent connections" (p. 13). Here, "the input layer is always determined entirely by external inputs and the output layer does not affect the rest of the system" (ibid.); hence a partition of the system's components into two subsets *A* and *B* with minimal or even zero values of EI($B \rightarrow A$) is possible.

So, what can we say about Φ in the case of a group agent? There are reasons to think that, in a typical group agent such as a corporation, court, or other organization, the value of Φ, while non-zero, would be low. Recall that Φ is low in a system if and only if it is theoretically possible to partition this system into two sub-systems such that the processes in one do not depend much on those in the other. In the human cerebral cortex, such a partition is not generally possible while keeping the functional architecture intact, and hence Φ is high. By contrast, I suggest that in a group agent, a "low-entropy partition" is theoretically (and sometimes even practically) possible. This is for at least three reasons.

Reason 1: 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. So, the group agent as a whole could not have a high value of Φ. At most, some smaller sub-units might do.

Reason 2: 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. Due to individually rational responses among the members and robust procedures, a modicum of orderly functioning would remain among the non-randomized rest of the unit. Hence the conditional entropy EI($B \rightarrow A$), where A and B are the two partition segments, would still be relatively low, suggesting a low level of Φ even for the "steering unit" of the group agent.

Reason 3: Much of the information processing in a group agent can be attributed to information processing by individuals. Arguably, the computational contribution made by cross-member connections, while non-negligible, is still moderate compared to the computational contribution made by individual cognitive processes. Schwitzgebel (2015, p. 1713) attributes a related point to David Chalmers:

> "Chalmers … has suggested (without endorsing) that the United States might lack consciousness because the complex cognitive capacities of the United States arise largely in virtue of the complex cognitive capacities of the people composing it and only to a small extent in virtue of the functional relationships between the people composing it."

I believe that the effect of all of this – given the formal definition of Φ – would be a low numerical value of Φ in a group agent, even if it is non-zero. It would not be even remotely close to the value of Φ that we would expect to find even in the brain of a

simple mammal, such as a rat or a mouse. And so, group agents would, at most, have a very small amount of consciousness according to integrated information theory.

It is worth noting that my argument does not depend on the claim that integrated information theory already *fully* captures the functional correlates of consciousness. The theory is still in its infancy, even if the idea of relating consciousness to informational integration turns out to be on the right track. For my argument against (non-negligible) group consciousness, it suffices to interpret integrated information theory as implying that a high value of Φ is a *necessary* condition for a correspondingly high level of phenomenal consciousness; it need not be *sufficient*.

Aaronson (2014), for instance, argues that informational integration alone is insufficient for consciousness. He suggests that there are some theoretically possible systems that have a very high value of Φ but that are not plausibly conscious, at least if we trust our intuitions. As evidence for this claim, he identifies a mathematically possible system, which he calls the "Vandermonde system", that performs a difficult but very mechanical number-crunching task and in doing so achieves a high numerical value of Φ. Yet, the system exhibits nothing like intentionality, agency, or genuine intelligence of the sort that we would ordinarily expect to find in a phenomenally conscious system.

It should be clear that even if a high numerical value of Φ is only necessary but not sufficient for phenomenal consciousness, my argument against group consciousness still stands. If I am right in thinking that Φ is low in a group agent, group consciousness could not really get off the ground.

6.4 Schwitzgebel's conclusion revisited

Interestingly, Schwitzgebel (2015) considers integrated information theory in his discussion of group consciousness but reaches a very different conclusion from mine. He argues that integrated information theory supports, rather than rules out, consciousness in a collective system such as the United States of America. He arrives at this conclusion by observing that a system such as the United States is highly functionally integrated, displaying "features like massively complex informational integration, functionally directed self-monitoring, and a long-standing history of sophisticated environmental responsiveness" (p. 1717). And he suggests that this implies that "the United States is at least a candidate for the literal possession of real psychological states, including consciousness" (ibid.). Indeed, the US's level of functional integration should compare favourably to that of a small mammal.

I agree with Schwitzgebel that the US is certainly a *candidate* for the possession of real psychological states. I also agree that, as noted earlier, the value of Φ for the US would be non-zero. At this point, one might follow Schwitzgebel in taking a "glass partly full" view about group consciousness. But I think that we are more warranted in taking a "glass largely empty" view. In particular, I think that, in ascribing "massively complex informational integration" to the US, Schwitzgebel employs a somewhat more informal notion of functional integration, albeit one that is well aligned with how functionalists would ordinarily understand that term. In the sense Schwitzgebel has in mind, the US is undoubtedly highly integrated. Yet, I believe

that, if we define Φ in the formal information-theoretic way described above, Reasons 1, 2, and 3 strongly count against a high value of Φ.

This is not to suggest that there could not be another, science-fiction kind of group agent with a higher value of Φ. But as things stand, the value of Φ in a typical real-world group agent is unlikely to be anywhere near the value we would expect to find in the kinds of brains we paradigmatically associate with consciousness.

6.5 The exclusion postulate

There is a further, and independent, reason why integrated information theory would speak against group consciousness. This reason is emphasized by Tononi and Koch (2015) in their argument that aggregates lack consciousness. It stems from one of the theory's central postulates: the *exclusion postulate*. According to it, whenever there is a nested hierarchy of sub-systems (such as parts of the brain nested in one another), consciousness is present at whichever layer maximizes informational integration, but not at any other layer.

So, although the brain in its entirety has a non-zero value of Φ, the cerebral cortex has a higher value of Φ, and hence the cortex, not the larger super-system, is the locus of consciousness. In a group agent, presumably Φ would peak at the level of the individual members, not at the collective level, and hence the members' individual consciousness would exclude the presence of consciousness at the collective level.

Generally, the exclusion postulate would rule out group consciousness, even if – contrary to what I have argued – the value of Φ were substantial for the group as a whole. The only condition would be that each individual member's value of Φ is still higher – which seems plausible, given the nature of the human brain. (Schwitzgebel 2015 acknowledges this argument but rejects the exclusion postulate on which it rests.)

Even if we do not accept this last argument, however, the earlier arguments, based on the low numerical value of Φ, should suffice to cast doubt on the existence of any significant amount of group consciousness.[19]

7. A normative implication

If my application of integrated information theory is correct, it supports the view that group agents have either very little or no phenomenal consciousness at the collective level (in line with the arguments in Tononi and Koch 2015). Of course, the argument is conditional on integrated information theory. However, among recent proposals concerning the functional correlates of consciousness, it is one of the more promising proposals. I would like to conclude by drawing attention to an important implication of the view I have defended.

In our book on group agency, Philip Pettit and I asserted a normative asymmetry between individuals and groups (List and Pettit, chs. 7 and 8). We argued that while

[19] My conclusion that there is group agency but no (non-negligible) group consciousness is consistent with Clark's conclusion, in relation to the "extended mind" hypothesis, that "[a]rguments for extended cognition … do not generalize to arguments for an extended conscious mind" (Clark 2009, p. 963).

group agents should be held responsible for their corporate actions, they should not be given the same rights as individuals, and they should be subject to especially strict checks and controls. So, a petroleum company, for example, should be held responsible for any oil spills it causes, but it should not be given the kinds of rights that individual human beings (ought to) enjoy. Rather, group agents should be given only those rights that can be defended within a normatively individualist framework. Such a framework is one that treats only individual people – and perhaps individual non-human animals – as ultimate units of moral significance, while assigning only derivative moral significance to collectives.

Thus any putative right of a group agent must ultimately be justified in terms of its contribution to the protection of individual rights and the promotion of individual interests. Some rights of group agents can be unproblematically justified in this way, such as the right to enter into appropriately regulated contractual relationships, for instance as an employer or as a participant in the market, or as a state making treaties with other states. But other rights are more problematic. Consider the controversy over how much freedom of expression corporations should enjoy.

In *Citizens United versus Federal Election Commission*, the US Supreme Court ruled that some First Amendment rights, such as free-speech rights, apply not only to individuals but also to certain corporate agents, so that the government cannot restrict the political-campaign contributions of corporate actors. The specific case was prompted by the question of whether, and to what extent, Citizens United, a conservative lobby group, had a free-speech right to publicly broadcast and advertise a film critical of Hillary Clinton during the 2008 presidential campaign (Liptak 2010).

Many commentators, including President Obama, criticized the Supreme Court's decision as detrimental to the good functioning of democracy. Here is what Obama said on the occasion of its fifth anniversary (as quoted in Alman 2015):

> "Our democracy works best when everyone's voice is heard, and no one's voice is drowned out. But five years ago, a Supreme Court ruling allowed big companies – including foreign corporations – to spend unlimited amounts of money to influence our elections. The Citizens United decision was wrong, and it has caused real harm to our democracy. With each new campaign season, this dark money floods our airwaves with more and more political ads that pull our politics into the gutter. It's time to reverse this trend. Rather than bolster the power of lobbyists and special interests, Washington should lift up the voices of ordinary Americans and protect their democratic right to determine the direction of the country that we love."

Evidently, Obama endorses the normative asymmetry between individuals and groups. But how can we philosophically justify it? Philip Pettit and I were criticized for not offering enough of a justification for it against the background of our defence of group agency and even group personhood (see, e.g., Briggs 2012). A key criticism was that the asymmetry seemed hard to defend in light of our functionalist account of agency. Why should we be normative individualists if we are collectivists about agency, where agency is defined in the same way for individuals and groups? Given the resources of functionalism alone, the answer to this question is not obvious.

The argument I have sketched in this paper offers a principled line of response. One could argue that an important difference in ultimate moral significance between individuals and groups lies precisely in their difference with respect to consciousness. Specifically, one might say that a *necessary* condition that an agent must satisfy in order to be of *non-derivative* moral significance is a capacity for phenomenal consciousness. Humans and other primates clearly have that capacity, while group agents do not – or at least not to any non-negligible extent.

On this picture, only agents that are of non-derivative moral significance – paradigmatically, individual human beings – can have non-derivative rights, or "rights in their own right". Other agents, such as group agents, can have, at most, derivative rights, which are of subordinate standing. We would then be justified in giving weaker rights to group agents than to individuals.

Citizens United and other corporations, so the argument goes, are less entitled to an unrestricted free-speech right than you or I are, because they are not conscious agents. And thus any rights that we might give to group agents would have to be justified in terms of their contribution to the protection of individual rights and interests. If giving a free-speech right to corporations fails to be in the interest of individuals, then no such right will be justified. Developing this point further is beyond the scope of this paper, but I hope to have said enough to put it on the table for discussion.

Let me close by returning to my original question: what is it like to be a group agent? Although my argument is tentative and conditional, it seems that the answer may well be: (close to) nothing.

References

Aaronson, S. (2014) "Why I Am Not An Integrated Information Theorist (or, The Unconscious Expander)", blog post at: http://www.scottaaronson.com/blog/?p=1799

Alman, A. (2015) "Barack Obama: 'The Citizens United Decision Was Wrong'", *The Huffington Post*, 21 January 2015; retrieved from: http://www.huffingtonpost.com/2015/01/21/barack-obama-citizens-united_n_6517520.html

Baars, B. J. (1988) *A Cognitive Theory of Consciousness*, Cambridge (Cambridge University Press).

Baars, B. J. (2003) "The global brainweb: An update on global workspace theory", Guest editorial, *Science and Consciousness Review*, October 2003.

Björnsson, G., and K. Hess (2015) "Corporate Crocodile Tears? On the Reactive Attitudes of Corporate Agents", manuscript.

Block, N. (1980) "Troubles with Functionalism?" In *Readings in Philosophy of Psychology*, Vol 1 (pp. 268-306), London (Methuen).

Block, N. (2008) "Phenomenal and Access Consciousness", *Proceedings of the Aristotelian Society* (New Series) 108: 289-317.

Bratman, M. E. (1999) *Faces of intention: Selected essays on intention and agency*, Cambridge (Cambridge University Press).

Bratman, M. E. (2014) *Shared agency: A planning theory of acting together*, Oxford (Oxford University Press).

Briggs, R. (2012) "The normative standing of group agents", *Episteme* 9(3): 283-291.

Carruthers, P. (2011) "Higher-Order Theories of Consciousness", *Stanford Encyclopedia of Philosophy*, available at: http://plato.stanford.edu/entries/consciousness-higher/

Chalmers, D. (1995) "Facing up to the problem of consciousness", *Journal of Consciousness Studies* 2(3): 200-219.

Chalmers, D. (1996) *The Conscious Mind*, New York (Oxford University Press).

Clark, A. (2009) "Spreading the Joy? Why the Machinery of Consciousness is (Probably) Still in the Head", *Mind* 118(472): 963-993.

Crick, F., and C. Koch (1990) "Towards a neurobiological theory of consciousness", *Seminars in the Neurosciences* 2: 263-275.

Crick, F., and C. Koch (2003) "A framework for consciousness", *Nature Neuroscience* 6(2): 119-126.

Dennett, D. (1987) *The intentional stance*, Cambridge, MA (MIT Press).

Dennett, D. (2005) *Sweet Dreams: Philosophical Obstacles to a Scientific Theory of Consciousness*, Cambridge, MA (MIT Press).

Dietrich, F., and C. List (forthcoming) "Mentalism versus behaviourism in economics: A philosophy-of-science perspective", *Economics and Philosophy*.

Fine, A. (1984) "The Natural Ontological Attitude", in J. Leplin (ed.), *Philosophy of Science* (pp. 261-277), Berkeley (University of California Press).

French, P. A. (1984) *Collective and corporate responsibility*. New York (Columbia University Press).

Gilbert, M. (1989) *On social facts*, New York (Routledge).

Goldman, A. (2004) "Group Knowledge Versus Group Rationality: Two Approaches to Social Epistemology", *Episteme* 1(1): 11-22.

Harnad, S. (2005) "Distributed Processes, Distributed Cognizers and Collaborative Cognition", *Pragmatics & Cognition* 13(3): 501-514.

Huebner, B. (2014) *Macrocognition: A Theory of Distributed Minds and Collective Intentionality*, Oxford and New York (Oxford University Press).

Huebner, B., M. Bruno, and H. Sarkissian (2010) "What Does the Nation of China Think About Phenomenal States?" Review of Philosophy and Psychology 1(2): 225-243.

Knobe, J., and J. Prinz (2007) "Intuitions about consciousness: Experimental studies", *Phenomenology and the Cognitive Sciences* 7: 67-83.

Levine, J. (1983) "Materialism and Qualia: The Explanatory Gap", *Pacific Philosophical Quarterly* 64: 354-361.

Liptak, A. (2010) "Justices, 5-4, Reject Corporate Spending Limit", *New York Times*, 22 January 2010.

List, C. (2014) "Free will, determinism, and the possibility of doing otherwise", *Nous* 48(1): 156-178.

List, C., and P. Pettit (2006) "Group agency and supervenience", *Southern Journal of Philosophy* 44(S1): 85-105.

List, C., and P. Pettit (2011) *Group Agency: The Possibility, Design, and Status of Corporate Agents*, Oxford (Oxford University Press).

Nagel, T. (1974) "What Is It Like to Be a Bat?" *Philosophical Review* 83(4): 435-450.

Oizumi, M., L. Albantakis, and G. Tononi (2014) "From the Phenomenology to the Mechanisms of Consciousness: Integrated Information Theory 3.0", *PLOS Computational Biology* 10(5).

Pettit, P. (2001) *A theory of freedom: From the psychology to the politics of agency*, Cambridge and New York (Polity and Oxford University Press).

Pettit, P. (2003) "Groups with minds of their own", in F. Schmitt (ed.), *Socializing metaphysics* (pp. 167-193), New York (Rowan and Littlefield).

Pettit, P., and D. Schweikard (2006) "Joint Action and Group Agency", *Philosophy of the Social Sciences* 36(1): 18-39.

Rovane, C. (1997) *The Bounds of Agency: An Essay in Revisionary Metaphysics*, Princeton, NJ (Princeton University Press).

Schwitzgebel, E. (2015) "If materialism is true, the United States is probably conscious", *Philosophical Studies* 172(7): 1697-1721.

Searle, J. R. (1995) *The construction of social reality*, New York (The Free Press).

Theiner, G. (2014) "A Beginner's Guide to Group Minds", in M. Sprevak and J. Kallestrup (eds.), *New Waves in the Philosophy of Mind* (pp. 301-322), London (Palgrave).

Tollefsen, D. P. (2002) "Collective intentionality and the social sciences", *Philosophy of the Social Sciences* 32(1): 25-50.

Tollefsen, D. P. (2015) *Groups as Agents*, Cambridge (Polity Press).

Tononi, G. (2015) "Integrated information theory", *Scholarpedia* 10(1): 4164.

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

Tuomela, R. (2007) *The philosophy of sociality: The shared point of view*, New York (Oxford University Press).

Tuomela, R. (2013) *Social Ontology: Collective Intentionality and Group Agents*, Oxford (Oxford University Press).