Introspection in Group Minds, Disunities of Consciousness, and Indiscrete Persons

Eric Schwitzgebel
Department of Philosophy
University of California, Riverside
Riverside CA 92521
USA

Sophie R. Nelson
Department of Philosophy
Oberlin College
Oberlin OH 44074
USA

June 29, 2023

Eric Schwitzgebel is Professor of Philosophy at University of California, Riverside. Among his books are *Describing Inner Experience? Proponent Meets Skeptic* (with Russell T. Hurlburt; MIT Press 2007), *Perplexities of Consciousness* (MIT Press 2011), and *The Weirdness of the World* (Princeton University Press, forthcoming).

Sophie R. Nelson is an undergraduate philosophy and psychology major at Oberlin College.

Introspection in Group Minds, Disunities of Consciousness, and Indiscrete Persons

Abstract: Kammerer and Frankish (this issue) challenge us to expand our conception of introspection beyond neurotypical human cases. This article describes a possible "ancillary mind" modeled on a system envisioned in Leckie's (2013) science fiction novel *Ancillary Justice*. The ancillary mind constitutes a borderline case between a communicating group of individuals and a single, spatially distributed mind. It occupies a gray zone with respect to personal identity and subject individuation, neither determinately one person or subject nor determinately many persons or subjects, and thus some of its processes might be neither determinately introspection within a mind nor determinately communication between minds. If ancillary minds defy discrete countability, the same might be true for some actual minds on Earth. Kammerer and Frankish's research program can be extended to include not only the study of possible forms of introspection, but also the study of possible mental activity intermediate between introspection and communication.

Word Count: ~5900 words

Keywords: consciousness, group minds, introspection, personal identity, science fiction

Introspection in Group Minds, Disunities of Consciousness, and Indiscrete Persons

Ordinarily, we think of persons as discrete, countable entities. A conference room contains zero people, or one, or two, or twenty-three, or some other whole number. If you ask how many people are present and someone answers "two and three-quarters", you'll probably think they're joking. (Maybe they're counting a fetus as a partial person.) If they answer "it's indeterminate between two and twenty-three", you'll presumably think that the indeterminacy concerns the criteria of presence (does attending via video call count as being "present"?) rather than concerning the discreteness of persons.

The discreteness of persons is, at best, a contingent fact about (most?) people as they happen to exist on Earth, not a fundamental fact about the nature of cognitively sophisticated systems, as will become evident once we consider how introspection might work in group entities. One consequence of the indiscreteness of persons is indeterminacy concerning the border between introspection within a mind and communication between minds, which implies a dimension of introspective variation beyond those described by Kammerer and Frankish (this issue).

1. Ancillary Minds: Perception, Action, and Memory.

Let's consider a cognitive system with what we'll call an *ancillary structure,* inspired by the cognitive architecture envisioned in Ann Leckie's (2013) science fiction novel *Ancillary Justice.* Orbiting a planet is a sophisticated computer. On the surface of the planet are ten platoons of twenty humanoid robots each. The orbiting computer and the humanoid robots are in constant communication via satellite. Without presupposing anything about the boundaries of persons, let's call the orbiting computer Aleph, and let's call each humanoid robot by platoon letter and member number, from A1 through J20. For purposes of this article, we will assume that broadly human-like cognition and consciousness is possible in computers and that functional and informational processes are what matter to consciousness. (If consciousness requires a brain, maybe we can instead imagine the robots to have modified organic brains with computer interfaces, which is actually closer to Leckie's original set-up.)

If all of the cognitive processing transpires in Aleph, while the robots are simply mobile input devices relaying raw camera and microphone feeds directly to Aleph, then the ancillary system presumably has one mind with many distributed input systems. If, in contrast, all of the cognitive processing transpires in the robots, and Aleph is just a relay system, then the ancillary system presumably has 200 minds, each in a separate location on the planet's surface. More interesting are the intermediate cases. (If you're wondering whether there might be both individual robot minds and *also* a group mind, hold that thought for now.)

Suppose, then, that substantial processing occurs in each robot before her signal is transmitted, and substantial processing also occurs in Aleph after the incoming signals are received, shaping Aleph's output signals to the robots. Here again, the limit cases are of secondary interest. If the signals have relatively thin content and each must be substantially processed before shaping the receiver's opinions and behavior, there are 201 minds (Aleph and the 200 robots, all separate individuals). If the signals are swift and thick, with direct access to cognitive and motor centers – so that it's approximately arbitrary whether some bit of processing occurs in a robot or in Aleph - then there's a single mind distributed across 201 locations. The target case is intermediate: The connectivity between the robots' cognitive centers and Aleph is swift, thick, and direct, but not so much that the ancillary entity constitutes a determinate case of a single, unified mind.

To better envision this possibility, consider some possible distributed perceptual, motor, and memory processes.

How might perception work? Suppose Platoon A is examining a field of flowers. Sixteen members of A, each employing their own local processing, initially represent the flowers as 50% blue and 50% red. Four members of A, in contrast, initially represent the same collection of flowers as 70% blue and 30% red. Aleph receives the conflicting signals. Maybe Aleph employs a winner-take-all procedure and represents the flowers as 50% blue. Maybe Aleph employs a weighted proportion procedure and represents the flowers as 50%*(16/20) + 70%*(4/20) = 54% blue. Maybe Aleph employs credences: a .8 credence in a 50/50 distribution and a .2 credence in a 70/30 distribution; or alternatively a .7 credence in 50/50, a .2 in 60/40, and .1 in 70/30. There are, of course, also many other, less simple alternatives.

Aleph's representation – suppose, for simplicity, that Aleph employs winner-take-all, representing the proportion as 50/50 – is then relayed to the platoons. Platoons B through J receive this as new information, having never seen the field. Sixteen members of A receive Aleph's signal as confirmatory of their local 50/50 representations. Four members of A receive Aleph's signal as disconfirmatory. They might then update their local representations to 50/50, discarding their initial representation. Alternatively, they might compromise between their initial representation and the incoming representation from Aleph: maybe 60/40, or maybe a 50% credence in 70/30 and a 50% credence in 50/50. Again, this just begins the possibilities. Further complexities can be introduced if some robots' representations deserve greater weight, due to their having better viewpoints or better equipment, or if the robots can communicate directly without Aleph as an intermediary.

How might action work? We can imagine, again, extreme cases: actions wholly controlled by Aleph, with the robots as passive limbs, or actions wholly controlled by the robots, with Aleph as a trusted advisor. But the interesting cases stand between. Maybe the local processes in Robot A1 form an action plan: Pick Flower 1. At the same time, an action plan arrives from Aleph: Pick Flower 2. Instead of assuming that the local plan always wins over the Aleph plan or vice versa, each action plan might carry a strength signal. If Pick Flower 1 has strength 3.4 and Pick Flower 2 has strength 2.6, then Pick Flower 1 will win, regardless of whether it originated in A1 or in Aleph, and the reverse if the strengths are reversed. A more complex model might compromise among ranked action plans: A1 ranks Flower 1 above Flower 3 above Flower 4 above Flower 2. Aleph ranks Flower 2 above Flower 3 above Flower 4 above Flower 1. Pick Flower 3 might then emerge as the best solution. Action selection might bee automatic and non-conscious, or it might involve conscious ambivalence, either in A1 or in Aleph or the ancillary entity as a whole, or in some blend of those locations. Once again, we have only gestured at some ways in which action planning might fail to be discretely localized. More complicated relationships are likely in a flexibly constructed ancillary agent.

How might memory work? For fast motor action, each robot might have substantial autonomy for basic procedural memories (e.g., how to open biosample jars) and short-term or iconic memory (e.g., of the exact position of the flowers as they sway in the wind). More complex skills might require complex interactions with Aleph, including longish feedback loops and some top-down guidance. Coordinated procedural memories – Robots A1 and A2 executing a specific, collaborative harvesting procedure, with A1 clipping the flowers and A2 bagging them – could be variously distributed among A1, A2, and Aleph. Semantic and episodic memories invite the same diversity of conflict resolution procedures as perception if the robots disagree. "External" memories might be stored as records in a cabin to which at least one robot has swift, reliable access; or they might be written on the surface of the robots' bodies, readable by other members of the robot's platoon; and Aleph might have no reliable means of knowing whether a memorial representation from Platoon A was retrieved from a cabin, body surface, or robot's internal cognitive store, processing memorial representations identically regardless of source (compare Clark and Chalmers 1998).

In writing and rereading this material we find it easy to slip into conceptualizing Aleph and the robots as distinct minds in communication; and when we try to resist that way of thinking, we find it easy to substitute the idea that the cognitive systems in Aleph and the robots are each subsystems of a single mind. The challenge we want to pose to you is to conceptualize the system in neither of those two ways or at least not *determinately* in either of those two ways, but rather as an architecture just unified enough to constitute a case that is borderline between communication within a mind and communication between minds.

For ease of exposition, we have assumed that if the number of minds is determinate it is either one or 201. However, other possibilities include there being 202 minds (one for each robot, one for Aleph, and one for the whole ancillary system), or 10 minds (one for each platoon), or a large number of partly overlapping minds (maybe one for each robot, one for each robot plus Aleph, one for each platoon, and one for each platoon plus Aleph). This of course further complicates the issue, creating further opportunities for gray zones and indeterminacy.

2. The Vague Boundary Between Introspection and Communication in Ancillary Minds.

We assume that introspection is a process by means of which one comes to represent one's own mind. Necessarily, if a process aims to represent someone else's mind, it is not introspective (Gertler 2000; Schwitzgebel 2010/2019; and probably Kammerer and Frankish this issue). From this it follows that if it is indeterminate whether the robots and Aleph are distinct minds or instead subparts of the same mind, it will sometimes be correspondingly indeterminate whether Aleph's queries to the robots constitute introspection within a mind or communication between minds. In an ancillary mind some mental activity might occupy an indeterminate gray zone between introspection and communication.

How might this work? For simplicity, we might imagine that Aleph has a local store of representations that she can access directly and that each robot similarly has a local store of representations that she can access directly. We emphasize the "for simplicity" qualification: Self-knowledge in human beings or in cognitive systems as sophisticated as the ancillary mind would likely tend to involve substantial "indirect" reconstructive and inferential elements (Schwitzgebel 2012; Kammerer and Frankish this issue). We bracket that issue to focus on possible introspection-like processes that could occur *between* Aleph and the robots.

In perception, as described above, Aleph monitors the perceptual judgments of the platoon members (A1 representing the flowers as 70% blue, A2 representing them as 70% blue, A3 representing them as 50% blue, etc.). It might be convenient for Aleph to continue to monitor the representational states of members of Platoon A. For example, suppose that A3 receives a corrective signal from Aleph which should, if A3 is functioning properly, change A3's representation of the field to 60% blue. Assuming that the robots' local representations contribute in some way to their behavior, A3's behavior will differ somewhat from the behavior of robots who represent the flowers as 50% blue. But maybe there's some noise in the corrective process, or maybe Aleph hasn't retained all the facts enabling that prediction, or maybe A3's representations are also corrected by signals directly from others in Platoon A, or maybe there's a chance of malfunction. The Aleph system might then compare her prediction about A3's representations with a confirmatory signal from A3.

In keeping with Kammerer and Frankish's (this issue) PID diagram, the process by which Aleph tracks A3's representations might be relatively direct – a simple triggering of an A3-represents-60% representation when A3 updates her perceptual representation to 60% – or the representation might involve a complex inferential procedure. It might be highly conceptualized – "A3 represents 60% blue flowers" – or it might be less conceptual, closer to analog, such as a detailed map of A3's perceptual states without high-level interpretive labels. It might be flexible – called on and implemented as needed, employing different types of inference and correction depending on the situation – or it might be inflexible and automatic, arising under prespecified conditions. If Aleph and the robots constitute an indeterminate number of minds, then these processes are indeterminate between communication and introspection. Kammerer and Frankish might thus consider adding another dimension to their portrayal of the varieties of introspection: a dimension corresponding to whether the process targets one's own mind (introspective) or another's mind (non-introspective), with a variety of intermediate cases nearer to or farther from canonically introspective or non-introspective processes.

If we assume that some configurations of the ancillary system generate the genuine, determinate existence of both a systemwide mind *and* individual minds for Aleph and each of the robots, another possibility appears: introspection at the systemwide level that is implemented by communication between minds at the individual robot level. For example, in response to a query from Aleph (an individual mind) the robots of Platoon A (also individual minds) might each communicate their representations of the flowers, and those responses might contribute to a systemwide representation, not only in Aleph but also in the cognition of the ancillary mind as a whole, of the form “75% of my Platoon A cognition represents the field as 70% blue and 25% of my Platoon A cognition represents the field as 60% blue.” Representing disagreement among its constitutive minds might then be an important part of the larger system’s introspective task. (Maybe such disagreement is felt as ambivalence.) Thus, a minded entity that contains other minds as parts might introspect its own mental states by means of verbal communication between the minds it contains.

3. The Indiscreteness of Some Ancillary Minds.

So far, we have assumed that the reader accepts that it is possible to create an ancillary system that has an indeterminate number of minds. That stipulation has, we think, some initial plausibility if one accepts that informational interconnectivity (of some sort or other) is the basis of mind individuation and comes in degrees. Connect two cognitive systems tightly enough (in the right way), and you have a single mind. Separate them sufficiently and you have two minds. Plausibly, one can construct a mind-merging sorites series (Roelofs 2019): Start with a paradigm case of two distinct minds. Connect them together slowly, bit by bit, until at the end stands a paradigm case of a single mind. Somewhere between, it seems, there must be borderline cases indeterminate between one mind and two. Pending some argument to the contrary, it would be surprising if there is a bright line at an exact spot, such that with one more bit of connectivity, any ancillary system would shift sharply, with no borderline cases, from being a 201-mind system to being a one-mind system (at least if one accepts a non-epistemicist, indeterminacy-allowing theory of vagueness).

This issue concerns not only *minds* but also *persons*, on the principle that, generally speaking, for cognitively sophisticated human-like systems, the number of minds equals the number of persons (though see Schechter 2018). One reason we do and should care about the boundaries between minds is that we care about the boundaries between persons: How many people die if the ancillary system is destroyed? If a promise comes from the mouth of A3, *who* is making the promise – just A3 or the ancillary system as a whole? If A3 throws herself into a volcano after instructions from Aleph, is that a person committing suicide at the command of another, or is it more like a single person choosing to sacrifice a limb? Our thinking about these questions about persons plausibly turns how we think about the individuation of minds, so that indeterminacy about mind individuation entails indeterminacy about person individuation.

If we accept the functionalist setup on which degree of connectivity is the basis of mind-and-person-individuation, sorites cases seem possible, and indeterminacy appears to follow. Furthermore, the type of indeterminacy is interestingly different from ordinary human cases of, say, a person slowly slipping into death, transitioning from one mind or one person to zero. In ancillary minds, the system might be determinately at least one person while being an indeterminate number of persons.

We see three main sources of resistance to this idea, which we'll call the Body View, the Phase Transition View (see also Schwitzgebel 2023), and Discrete Phenomenal Realism. According to the Body View, we count up persons by counting up bodies. In the ancillary mind, each robot has one body. Aleph has a body (arguably). Therefore, 201 persons exist, regardless of the cognitive connections between them. Note that the Body View might permit indiscrete persons the number of bodies can be indeterminate. But we won't address the Body View further. Instead, we will just assume its falsity; our interest is in persons as cognitive entities.

Gaps merely of space and form, crossed without meaningful delay due to the speed of light, are irrelevant to person-individuation in our intended sense (see also Dennett 1978; Parfit 1984; Schwitzgebel 2015).

The Phase Transition View observes that sometimes sudden phase transitions disrupt what might seem to be continua. Water cools and cools, not changing much, then suddenly at 0.0° C it becomes ice. A beam strains and strains under increasing weight, bending a bit more with each added gram, then suddenly it snaps. After the transition, much is different. You can rest a nickel on the ice. You can wiggle one end of the beam without wiggling the other. Similarly, it is possible at least in principle that something similar could happen in the ancillary system. Integrate it slowly, a bit at a time, and then suddenly, at some point, with some tiny bit more integration, it changes radically. Processes that would have worked one way now work quite another way. The processes that constitute perception, memory, action, and introspection shift all in a twinkling from following one type of pattern, the 201-person pattern, to following a very different pattern, the one-person pattern. To count as a proper phase transition, this would have to be a substantial shift in the functioning of the system overall, not just a subtle change, like the shift from one shade of green to a barely different shade of not-quite-greent. Only a sudden and substantial change creates a meaningfully sharp boundary between the functional operations of the one-person case and the 201-person case, rather than an arbitrary or semi-arbitrary different labeling of neighboring cases. With a tiny bit more integration, the system must tip into a very different local minimum, like the tiniest nudge of a basketball on rim's edge tips it into rather than out of the hoop, with all the different causal consequences that follow.

The Phase Transition View is not strictly impossible. However, there appears to be no reason to suppose that there would always be such sudden functional phase transitions in the course of integration, much less that whatever phase transitions there are should align exactly with the proper boundary between the one-person and the 201-person case, and still less that *all possible ways* of building an ancillary system would have sharp phase transitions in exactly the right places. Furthermore, if an advocate of the Phase Transition View holds that there can never, even for an instant, be indeterminate cases, they face the problem that sharp transitions admitting of no borderline cases whatsoever arguably are rare among macroscopic entities in nature. For these reasons, pending further argument, the Phase Transition view seems empirically implausible.

One might avoid that empirical risk by accepting the possibility of a sharp metaphysical change atop a smooth gradation of functionally similar cases, so that neighboring cases in a sorites series that differ perhaps only by a single bit of connectivity – with no big downstream functional differences– can nonetheless be sharply and determinately different in the number of persons. Case N is sharply and determinately 201 persons; case N+1, empirically almost indistinguishable, is determinately one person, with no cases between.

Perhaps the most attractive version of this view is Discrete Phenomenal Realism, according to which, atop all of the representational, informational, cognitive processes described in Sections 1 and 2, exist some simple phenomenal facts concerning personhood or subjectivity. These phenomenal facts might be further, nonphysical facts (as in property dualism) or they might be physical facts grounded in the ancillary's physical systems. This view draws its plausibility, at least in part, from the attractiveness of treating phenomenality, subjectivity, or "what-it's-like"-ness as necessarily discrete, without indeterminate, borderline cases (Antony 2008; Goff 2013; Simon 2017; though see Schwitzgebel 2023 for a counterargument). Either there's something it's like to be Robot A1, and Robot A2, and Robot A3, and so forth, generating a total of 201 systems of which it's correct to say there's something it's like to be them – in which case there are 201 persons or at least 201 subjective centers of experience – or there's nothing it's like to be these systems, and instead there's something it's like to be the ancillary system as a whole – in which case there is only one person or subjective center of experience. Either there's one thing it's like to be the whole, or there are many things it's like it be the parts. "Something-it's-like-ness" can't occur an indeterminate number of times. Phenomenality or subjectivity must have sharp edges, the thinking goes, even if the corresponding functional processes are smoothly graded.

Discrete Phenomenal Realism also fits nicely with Bayne's (2010) account of the unity of consciousness and with "phenomenalist" approaches to introspection of the sort described by Kammerer and Frankish (this issue). Imagine one locus of consciousness in which experiential states α, β, γ, and δ occur, such that each state is introspectible by the same subject, Subject 1. A second locus of consciousness hosts experiential states ε, ζ, η, and θ, all introspectible by Subject 2. A third locus hosts ι, λ, μ, and ν, introspectible by Subject 3. No subject can introspect any of the states that are introspectible by any other subject, and there are no free-floating experiential states had by no subject. Phenomenal "co-occurrence" relations are similarly neat: There's a unified state in which α, β, γ, and δ all occur together (in the way that your simultaneous experiences of the sound of music, taste of beer, and view of the band all phenomenologically co-occur at a concert); another unified state in which ε, ζ, η, and θ co-occur (as your friend's similar experiences of the concert co-occur in your friend but not in you); a third in which ι, λ, μ, and ν co-occur; and no co-occurrence relationships of any other sort (e.g., of α, ε, and ι). Count up the introspective subjects or loci of mutual phenomenal co-occurrence. That's your number of persons. The result will always be a whole number.

Discrete Phenomenal Realism can't be right if, as Kammerer and Frankish maintain, illusionism about phenomenal consciousness is true. Discrete phenomenal realism presumably requires phenomenal realism. But even for non-illusionists, the view faces two problems.

First, it inelegantly posits sharp ontological lines atop fuzzy underlying continua. If the Discrete Phenomenal Realist rejects the Phase Transition view, then ancillary systems can be very similar to each other in all representational, informational, and behavioral respects, yet radically different in their introspective and phenomenal co-occurrence relationships – one system hosting 201 separate streams of experience, for example, and the other hosting only a single stream. Given their representational, informational, and behavioral similarity, these systems will seem virtually the same from outside, and the robots, Aleph, and the system as a whole will give virtually the same outputs when queried about their mental states. There will be no phase-shift-like announcement of an experience of sudden unity, and there will be no "Ahhh! I've split into 201 pieces! Help, help!" Add or subtract a smidgen of connectivity, and the system leaps across the gulf from one to 201 minds and back again with no significant accompanying psychological change. Thus, the view strangely dissociates consciousness from psychology (compare Frankish's 2021 critique of panpsychism).

Second, the Discrete Phenomenal Realist will struggle to accommodate cases of apparent partial experiential overlap. Modifying the ancillary structure somewhat, suppose that Robot A1 and Robot A2 share much of their circuitry in common. Between them hovers a box in which most of their cognition transpires. Maybe the box is connected by high-speed cables to each of the bodies, or maybe instead the information flows through high bandwidth radio connections. Either way, the cognitive processes in the hovering box are tightly cognitively integrated with A1's and A2's bodies and the remainders of their minds – as tightly connected as is ordinarily the case in ordinary unified minds. Despite the bulk of their cognition transpiring in the box, some cognition also transpires in each robot's individual body and is not shared by the other robot. Suppose, then, that A1 has an experience with qualitative character α (grounded in A1's local processors), plus experiences with qualitative characters β, γ, and δ (grounded in the box), while A2 has experiences with qualitative characters β, γ, and δ (grounded in the box), plus an experience with qualitative character ε (grounded in A2's local processors). If indeterminacy concerning the number of minds is possible, perhaps this isn't a system with a whole number of minds. However, the Discrete Phenomenal Realist must attribute a determinate number of minds, and will need to make sense of the case in a different way.

As we see it, Discrete Phenomenal Realists have three options: Impossibility, Sharing, and Similarity. According to Impossibility, the setup is impossible. However, it's unclear why such a setup should be impossible, so pending further argument we disregard this option. According to Sharing, two determinately different minds share tokens of the very same experiences with qualitative characters β, γ, and δ. According to Similarity, there are two determinately different minds who share experiences with qualitative characters β, γ, and δ but not the very same experience tokens: A1's experiences β_1, γ_1, and δ_1 are qualitatively but not quantitatively identical to A2's experiences β_2, γ_2, and δ_2. An initial challenge for Sharing is its violation of the standard view that phenomenal co-occurrence relationships are transitive (so that if α and β phenomenally co-occur, and β and ε phenomenally co-occur, so also do α and ε). An initial challenge for Similarity is the peculiar doubling of experience tokens: Because the box is connected to both A1 and A2, the processes that give rise to β, γ, and δ each give rise to two instances of each of those experience types, whereas the same processes would presumably give rise to only one instance if the box was connected only to A1.

To make things more challenging for the Discrete Phenomenal Realist who wants to accept Sharing or Similarity, imagine that there's a switch that will turn off the processes in A1 and A2 that give rise to experiences α and ε, resulting in A1's and A2's total phenomenal experience having an identical qualitative character. Flipping the switch will either collapse A1 and A2 to one mind, or it will not. This leads to a dilemma for both Sharing and Similarity.

If the defender of Sharing holds that the minds collapse, then they must allow that a relatively small change in the phenomenal field can result in a radical reconfiguration of the number of minds. The point can be made more dramatic by increasing the number of experiences in the box and the number of robots connected to the box. Suppose that the 200 robots each have 999,999 experiences arising from the shared box, and just one experience that's qualitatively unique and localized – perhaps a barely noticeable circle in the left visual periphery for A1, a barely noticeable square in the right visual periphery for A2, etc. If a prankster were to flip the switch back and forth repeatedly, on the collapse version of Sharing the system would shift back and forth from being 200 minds to one, with almost no difference in the phenomenology. If, however, the defender of Sharing holds that the minds don't collapse, then they must allow that multiple distinct minds could have the very same token-identical experiences grounded in the very same cognitive processors. The view raises the question of the ontological basis of the individuation of the minds; on some conceptions of subjecthood, the view might not even be coherent. It appears to posit subjects with metaphysical differences but not phenomenological ones, contrary to the general spirit of phenomenal realism about minds.

The defender of Similarity faces analogous problems. If they hold the number of minds collapses to one, then, like the defender of Sharing, they must allow that a relatively small change in the phenomenal field can result in a radical change in the number of minds.

Furthermore, they must allow that distinct, merely type-identical experiences somehow become one and the same when a switch is flipped that barely changes the system's phenomenology. But if they hold that there's no collapse, then they face the awkward possibility of multiple distinct minds with qualitatively identical but numerically distinct experiences arising from the same cognitive processors. This appears to be ontologically unparsimonious phenomenal inflation.

To review: If we accept that the individuation of minds and persons is grounded in facts about functional interconnection that can in principle be arranged in a sorites series with indeterminacy in the middle, then we're forced to one of two options: Either accept that the number of minds and persons can be indeterminate, or insist on a metaphysical bright line despite the hypothetical constructability of a sorites series. The second option divides into two sub-options: If there is a metaphysical bright line, it might or might not involve a big functional difference. If it does involve a big functional difference (the Phase Transition View), then intermediate configurations in the sorites series must be functionally unstable, collapsing quickly into functionally distinct, discretely countable mind/person configurations. There seems to be little reason to think, empirically, that this must be so. If it doesn't involve a big functional difference, then it looks like the best option is Discrete Phenomenal Realism. But Discrete Phenomenal Realism seems to require oddly bright metaphysical lines in the absence of big functional differences. Moreover, different variations of Discrete Phenomenal Realism all appear to face serious problems accounting for the possibility of qualitative similarity between minds.

4. Back to Earth: Countries, Multiple Personality, Split-Brain Cases, and Craniopagus Twins.

If we're willing to abandon discrete countability and allow introspection-like mental processes for far-out science fiction cases, then maybe we should do so for some cases on Earth.

In other work (Schwitzgebel 2015), one of us has argued that it is not unreasonable to think that the United States, conceived of as a concrete entity with people as its parts, literally possesses a stream of consciousness over and above the experiences of the citizens and residents that constitute the United States. This entity represents, metarepresents, engages in massive, complex information processing, and responds in intelligent ways to its environment. One concern for this view is that if we accept that the United States is conscious, we might also have to accept that Luxembourg and the Microsoft Corporation are conscious; and if we accept that Luxembourg and the Microsoft Corporation are conscious, we might also have to accept that Delaware and U.C. Riverside are conscious; and so on, down to a position even more seemingly absurd than the idea that the United States is conscious. If consciousness is not an all-or-nothing matter and conscious entities need not always be discrete and countable, then perhaps consciousness fades gradually rather than suddenly as group entities become less like a single, interconnected mind. If conscious minds can be present but indiscrete in hypothetical ancillary systems, then perhaps they can be present but indiscrete in actually existing groups on Earth.

Some human cases might also involve indeterminacy regarding the number of minds and persons, as well as in the presence of introspection versus communication. Well-known problem cases include Dissociative Identity Disorder or Multiple Personality (Braude 1995; Radden 1996; Maiese 2016), split-brain cases (Schechter 2018), craniopagus twins with overlapping brain regions (Cochrane 2021), and maybe some non-human animals, such as the octopus, with substantially disunified cognitive processing (Godfrey-Smith 2016). Faced with the Hogan twins, or with Jekyll-Hyde, the correct answer to the number of minds or persons might be neither one nor two but something indeterminate between. The Hogan Twins and Jekyll-Hyde might sometimes engage in cognition that is neither determinately introspection within one mind nor determinately communication between two minds. If we accept that, then perhaps even typical minds are less discretely individuated than we tend to think. Maybe we’re all a bit fuzzy-bordered, disunified, and plural.[1]

[1] Thanks to Rotem Herrmann concerning memory in an ancillary system (see also Schwitzgebel and Herrmann 2016) and Nichi Yes concerning personal identity in Dissociative Identity. For useful comments on the draft, thanks to Brice Bantegnie, Tim Bayne, Rotem Herrmann, François Kammerer, and Luke Roelofs.

References

Antony, Michael V. (2006). Vagueness and the metaphysics of consciousness. *Philosophical Studies, 128*, 515-538.

Bayne, Tim (2010). *The unity of consciousness.* Oxford University Press.

Braude, Stephen E. (1995). *First person plural.* Rowman & Littlefield.

Clark, Andy, and David Chalmers (1998). The extended mind. *Analysis, 58,* 7-19.

Cochrane, Tom (2021). A case of shared consciousness. *Synthese, 199,* 1019-1037.

Dennett, Daniel C. (1978). Where am I? In D.C. Dennett, *Brainstorms,* MIT Press.

Frankish, Keith (2021). Panpsychism and the depsychologization of consciousness. *Aristotelian Society Supplementary Volume, 95,* 51-70.

Gertler, Brie (2000). The mechanics of self-knowledge. *Philosophical Topics, 28* (2), 125-146.

Godfrey-Smith, Peter (2016). *Other minds.* Farrar, Straus and Giroux.

Goff, Philip (2013). Orthodox property dualism + the Linguistic Theory of Vagueness = Panpsychism. In R. Brown, ed., *Consciousness inside and out.* Springer.

Leckie, Ann (2013). *Ancillary justice.* Orbit.

Maiese, Michelle (2016). *Embodied selves and divided minds.* Oxford University Press.

Parfit, Derek (1984). *Reasons and persons.* Oxford University Press.

Radden, Jennifer (1996). *Divided minds and successive selves.* MIT Press.

Roelofs, Luke (2019). *Combining minds.* Oxford University Press.

Schechter, Elizabeth (2018). *Self-consciousness and "split" brains.* Oxford University Press.

Schwitzgebel, Eric (2010/2019). Introspection. *Stanford Encyclopedia of Philosophy* (winter 2019 edition).

Schwitzgebel, Eric (2012). Introspection, what? In. D. Smithies and D. Stoljar, eds., *Introspection and consciousness.* Oxford University Press.

Schwitzgebel, Eric (2015). If materialism is true, the United States is probably consciousness. *Philosophical Studies*, *172,* 1697-1721.

Schwitzgebel, Eric (2023). Borderline consciousness, when it's neither determinately true nor determinately false that experience is present. Unpublished manuscript at http://faculty.ucr.edu/~eschwitz/SchwitzAbs/BorderlineConsciousness.htm.

Schwitzgebel, Eric, and Rotem Herrmann (2016). Possible architectures of group minds: Memory. Blog post at *The Splintered Mind* (Jun 14). https://schwitzsplinters.blogspot.com/2016/06/possible-architectures-of-group-minds.html.

Simon, Jonathan A. (2017). Vagueness and zombies: Why 'phenomenally conscious' has no borderline cases. *Philosophical Studies*, *174,* 2105–2123.