Studying the whole of consciousness

Colin Klein and Andrew Barron

The mystery of consciousness, at its core, stems from the relationship of the whole to its parts. The raw material of consciousness is individual sensations. Yet conscious sensations are only conscious *for* and *to* someone: that is, for a whole locus of consciousness, which is in turn constituted by its parts.

Most of us start with the whole, and encounter the mystery by contemplating others. Sometimes it starts with difference: can the teacher see what I can't see? How can he find cricket so thrilling? Often, it's through awe at the sheer scope of human experience: What was it *like* for my father-in law-when he was a young refugee, trekking back in forth along the length of the Korean Peninsula? Or for Reinhold Messner, winded and triumphant, standing atop Everest?

These are hard questions, even though they concern other humans – humans with sensory systems roughly like our own, doing things roughly like the things you do. I can sharpen the problem even further by considering what it was like to be *me* when I was younger. Olivia Bailey (Bailey, 2023), echoing Hume, introduces the useful concept of a *sensibility*: "an emotional orientation to the world" that "shapes how things appear to us."

My aesthetic taste is one kind of sensibility. At the museum I see not just simple colours and shapes but whole artworks, saturated with feelings and reactions. That entire orientation to the work shapes my attention and my emotional engagement with the piece. Yet tastes change. As Bailey notes:

> "Looking at Fragonard's *The Swing*, I wonder: how could I have ever felt that frivolous nonsense was the height of beauty and refinement? *The Swing* has not changed. It is still composed of the same brushstrokes, in the same arrangement. I can see those features now, I know they had a beautiful appearance, and I have some memory of what it was like to find them beautiful. Still, I cannot now recall, recreate, or otherwise revisit that beautiful appearance. To put it another way, I cannot now see or picture *The Swing* as beautiful, and so I cannot find my old admiration for it intelligible. (Bailey, 2023, p188).

Bailey's feeling of unintelligibility is directed *at herself*: her change in aesthetic sensibility as she matured makes her younger experience feel alien. Such changes are hardly unusual:

I also know what it is like to age, and to realise that my aesthetic, intellectual, ethical, emotional or physical sensibilities have shifted – rendering my younger self not merely distant but faintly *confusing*. I cannot quite remember what it was *like* to be moved in that way, or to fail to appreciate the pieces that I now cherish.

Bailey's example emphasises the degree to which the mystery of consciousness is not *just* the mystery of raw sensation. Sensations are odd, of course. To make the mystery of consciousness more vivid, philosophers have often emphasised alien sensations: the bat's echolocation (Farrell, 1950; Nagel, 1974), or unseen colours (Jackson, 1986), or the taste of vegemite (Lewis, 1990) and durian (Paul, 2016).

Yet exotica are unnecessary and, perhaps, misleading. It is a commonplace that novel tastes are hard to describe. We've been told since childhood that there are new things to come that we must experience to know (Farrell, 1950). What Bailey finds so odd about revisiting Fragonard is that there is an obvious sense in which she sees the *very same thing* that now strike her differently.

What it is to have your perspective on the world – not in the simple sense of what you see from where you stand, but in the achingly particular way the world seems to you – constitutes a kind of *global* sensibility. It is the way that each sensation is infused with meaning through its links to the very particular life you've lived, the body you inhabit, and the things you care about. To understand another person, even a younger you, seems like it would involve taking up a different sensibility. But, to foreshadow a bit, this has the faint whiff of paradox: it seems impossible because the subject who managed it would no longer to be *you*. Yet sometimes, I *do* manage to understand what it is like to be someone else. That is important too.

We think that the mystery of consciousness does not arise (as many philosophers suppose) from the gulf between objective and subjective. That gap is the symptom, not the cause. Instead, we'll suggest, the mystery stems from the interplay between local and global – between individual sensations and the whole feeling subject that has them. Examining the link between global and local gives us a better picture of how consciousness ought to be studied, suggests why conscious minds evolved in the first place, and gives a picture of where the science of consciousness ought to go next.

♦♦♦

Military historian John Keegan, teaching Sandhurst cadets about the psychology of battle, begins not with what was seen or believed but with what was *felt*:

> It is a fairly safe generalization that the soldiers of most armies, at least before the development of mechanical transport, entered battle tired, if only because they had had to march to the field under the weight of their weapons and kit. The English army at Agincourt was certainly very tired, and hungry, cold and wet into the bargain (Keegan, 2011, p129).

To understand the English archer, we must start with his bodily sensations – hunger, thirst, pain, fatigue, and thermal discomfort. It is a good place to start thinking about consciousness, too. Bodily sensations have the best claim to being common among all organisms who feel anything at all: to be alive is to have a body that needs protecting.

That is why Aristotle in the *De Anima* singled out touch as the first and most indispensable sense. All animals, he says "have at least one sense, touch: and, where sensation is found, there is pleasure and pain, and that which causes pleasure and pain; and, where these are, there also is desire, desire being appetite for what is pleasurable" (DA 414b1-4). Touch is thus intimately wrapped up with appetite and desire, and with appetite comes pleasure and pain. All of this, says Aristotle, arises from the demands of self-motion. For when animals could move, they needed to know where to move *to* – and that requires knowing what you need and what you'd prefer to avoid. Hence hunger, thirst, pain, cold: all as a package.

Yet even here, one can't simply assume that bodily sensations can simply be transposed from one person to another. Keegan doesn't. The hunger and fatigue of the archer at Agincourt were felt not by a modern man writing at a comfortable kitchen table but by an illiterate peasant, for whom war was at one end of a spectrum of daily violence, and who even if successful had a very long walk home. To feel is always to have that feeling *mean* something to you.

Thomas Nagel emphasised the challenge of trying to understand another's sensibility in his famous essay "What is it like to be a bat" (1974). To return to a philosophical register: the phrase 'what it is like' is fundamentally *relational*. It is (roughly speaking) about the link between an experience and the subject that has the experience. In a thoughtful philosophical piece, Daniel Stoljar analyses 'what it is like' statements as postulating what he calls *affective relations*. An affective relation holds between individuals and events "just

in case the individual is affected in a particular kind of way by the event" (Stoljar, 2016, p1162). Analysing 'what it is like' in terms of conscious relations makes clear that conscious states are always, in Nagel's term "subject-involving' (Nagel, 1974, p1191).

So when I try to understand what it's like even to feel a particular experience, there are two ends to the *relata*: the experience, and the subject of experience. This means that the point of Nagel's thought experiment is not (just) to show that unaided imagination can fail to capture novel sensations. (The philosophical significance of this was always a bit dubious anyway). Nagel recognised this. He points out that he could try to imagine hanging upside down in the rafters and eating insects, but "Insofar as I can imagine this (which is not very far), it tells me only what it would be like for *me* to behave as a bat behaves. That is not the question. I want to know what it is like for a *bat* to be a bat." (Nagel, 1974) The italics emphasise a whole subject of experience.

Return to pain and hunger. One good reason to focus on bodily sensations is that bodily sensations are *obviously* subject-involving. The idea of a free-floating pain or a hunger had by nobody is faintly ridiculous. Hunger and pain are imperatives (Klein, 2015): commands felt as issued by your body and that demand kinds of action. Hunger says *eat!*, pain *protect yourself!*, and so on. Commands always have an issuer and a recipient (Hamblin, 1987), so the fact that it is you being directed is always front and centre.

(An important aside: while we focus on the drives that support bodily integrity, the drive to reproduce is at least as important. From the point of view of evolution, an organism that leaves no successors is not materially different from one that never existed at all. Given that reproduction is often at odds with survival, deciding what to do gets trickier still.)

One can, of course, talk about the objects of consciousness without mentioning a subject. Much philosophy does. The standard philosophical examples of odd tastes and novel colours let the subject fade into the background: the weirdness of vegemite comes from its strange combination of salt, umami, and malt. Saying it that way focuses on the world, letting the subject recede. Colours are worse still – it's hard to see why *this* sensation ought to go with grass, that with the sky.

Talking thus gives the misimpression of conscious states as simple, *sui generis* properties. It can all end up seeming a bit arbitrary, why consciousness is the way it is, or, to be honest, why you are conscious of anything in the first place. I don't forget myself when it comes to bodily sensations. Hunger is always *my* hunger. That is why it motivates me. So does

my thirst, and my fatigue, and my pain. Worse, I often have to sort out different, incompatible bodily commands. There's only a conflict because there's one thing, me, that's the subject of all of them, and I can't fix everything at once. Indeed, the possibility of this sort of conflict may be the key to understanding why we are conscious in the first place.

♦♦♦

As we write, I am watching a wattle bird feed on the banksia flowers in the bushes just beyond the balcony. It's a rainy winter morning in Australia. The bird is wet, chilled, hungry, thirsty, and fatigued. Other animals threaten: fellow wattle birds compete for its flowers, the wedge-tailed eagle circles high overhead, the vizsla wags its tail as it stares with rapt attention.

Each action the wattle bird takes resolves these competing demands, in a moment-by-moment way. I am no different, even if my needs can be more abstract: to pause from writing to stretch, make another cup of coffee, and watch the wattle bird is *also* a resolution of my competing imperatives.

The active choice of what to do now is known as the *action selection problem*. For complex living things the action selection problem can be deceptively, fiendishly complex. For starters, different simultaneous imperatives are resolved by different, often incompatible actions. Foraging requires the wattle bird to expend energy and risk exposure to cold and threats. Staying in a tree hollow is to say safe and protected, but eventually starvation threatens. Some threats are certain (foraging in the rain means wet and cold). Some are uncertain (predators are crafty). So too with rewards: foraging for food is also uncertain, and the bird may have to travel some distance with no guarantee of success. To top it off, effective action requires commitment: half-hearted foraging leaves you cold *and* hungry.

How to resolve this space of incompatible imperatives, rewards and risks? Mathematically speaking, there is no easy solution to the action selection problem, and there are often no computationally tractable exact solutions (Hayes et al., 2022; Roijers, Vamplew, Whiteson, & Dazeley, 2013). For starters, the problem usually involves a number of difficult tradeoffs. The value of different resources or costs are relative to the state of the bird. If the bird is producing eggs the relative value of protein is far higher than carbohydrate; for the empty-nester this relationship flips around. If the bird is brooding a clutch of eggs then the relative

cost of leaving the nest is far greater than at other times, since exposed eggs cool quickly. The relationship is also nonlinear: eating and returning to the nest is very good, while failing either to eat or to return is quite bad.

The action selection problem is also an irreducibly *multi-dimensional problem*. While economists thought that one could convert all good things to a single scale of utility, there's good reason to think that this is biologically unrealistic (Niv, Joel, Meilijson, & Ruppin, 2002). Intuitively speaking, it is very hard to give a direct comparison between the value of another cup of coffee and incrementally advancing towards a deadline. And of course, it's harder still when uncertainty is involved, especially because the uncertainty involved is different for different actions. Maybe the coffee will be especially good, maybe the grant proposal will get through: the *kind* of evidence that is relevant to both is different.

Finally, a solution is not optional. For in the end, two stark facts govern all biological needs. One: you have only one body with which to act. Bjorn Merker, (2007) from whom we take much inspiration, calls this the 'final common pathway'. Two: that no need can be put aside forever: that the failure of one need to be satisfied leads to death, which is the failure of all needs.

Nevertheless, the bird solves it. So do I. So do many animals. How?

Start with what a satisfactory solution would look like. Action selection requires calculating pairs of actions and expected values. So a solution has bring together diverse sources of information into a common relational framework. This must incorporate facts from the senses about the state of the external world, representations of internal physiological state, and learned knowledge of value and the past utility of resources. This integration must be global, in the specific sense that it allows different pairs of options to be meaningfully contrasted.

We have elsewhere referred to the mechanism that solves this problem (Klein & Barron, In press) as a *phenomenal interface*. *Interface* because it brings together multiple diverse streams of information into a common relational framework. *Phenomenal* because computations of this kind have structural properties that are relevant to our conscious experience of our sensibilities. It is not accidental that our conscious bodily sensations are experienced as part of a global sensibility: the *point* of consciousness is to create such a sensibility. Sensory stimulation only becomes conscious when it is processed as part of a phenomenal interface.

By way of illustration, return again to bodily sensations. Most bodily sensations are actually pretty obscure about their causes (most of our aches and pains have unclear origins, and I get hungry throughout the day despite the fact that I am far from starvation). Pains are conscious not to inform but as key inputs to a process of adjudicating what comes next. Hunger and thirst only have meaning when they are placed in relation to what could be done to satisfy them and how that relates to other demands. It is pointless to compare hunger and thirst in isolation. If I am both hungry and thirsty I will forgo distant food if water is close at hand. Our experience of bodily sensations depends not just on a bodily imperative but also on the opportunities to resolve that imperative. Thirst intensifies when a drink is in hand, and is slaked immediately after drinking, even though there is no time for consumed water to have entered the blood stream and restored homeostasis. The need to urinate is easily ignored at a bar with friends or on the train ride home, but becomes utterly overwhelming when fumbling with keys for the front door.

Consciousness is the construction of meaningful sensations, but it doesn't show its work. Often we can't pin down accurately exactly why we make any given choice – why break for a coffee now rather than in ten minutes time? Similarly with feelings: you might have no idea why you're hungry, or angry; what's vivid is only a demand for action. An integrated phenomenal interface means we cannot access the contributions of any one *reason* for a decision. We experience the outcome, but we have limited access under the hood at the mechanism of the decision. This is partly why consciousness itself feels so mysterious, and why conscious sensations can sometimes feel quite arbitrary. What is relevant is not any particular feeling and its reasons, but its importance for the whole agent experiencing it.

♦♦♦

So consciousness solves a fundamental (and thorny) problem – deciding what to do next. Does this mean that all life is conscious? We don't think so. Once it is clear what consciousness does, it is also clear what doesn't need it.

In his breathless panegyric to the nematodes – nearly invisible, often parasitic worms – Cobb notes their incredible abundance:

> …if all the matter in the universe except the nematodes were swept away, our world would still be dimly recognizable, and if, as disembodied spirits, we could then

> investigate it, we should find its mountains, hills, vales, rivers, lakes, and oceans represented by a film of nematodes. The location of towns would be decipherable, since for every massing of human beings there would be a corresponding massing of certain nematodes. Trees would still stand in ghostly rows representing our streets and highways. The location of the various plants and animals would still be decipherable, and, had we sufficient knowledge, in many cases even their species could be determined by an examination of their erstwhile nematode parasites. (Cobb, 1914)

Nematodes are alive. They have complex needs. They have brains, often large for their size. The most-studied nematode, *Caenorhabditis elegans* (*C. elegans*), has 302 neurons, nearly a third of its cells. Yet nematode consciousness does not form a ghostly shroud across the world. Why? Because nematodes aren't conscious. Neither are earthworms, bacteria, slime moulds, plants, or mushrooms. They solve their action selection problems, but don't need a phenomenal interface.

C. Elegans is instructive in this regard. It is surprisingly complex, and has an exquisitely studied nervous system and brain. It has the honour of being the first animal to have its nervous system completely mapped. It has multiple senses, and especially acute chemosensation. It becomes more sensitive to food stimuli when hungry. Yet *C. elegans* can solve the action selection problem relatively simply. What it pursues at a time is mostly a linear function of what it lacks. It doesn't need to plan ahead, because the only relevant information it has about its world is what is impacting its body wall at any time. So it can solve its action selection problem without the need for a phenomenal interface just by setting different motivations (each gated by relevant sensory information) in opposition with each other, and the strongest wins.

Similar remarks apply to jellyfish, plants, trees, and fungi. All are complex, and all make adaptive changes to their world in response to stimuli. But there is no need for the sort of difficult global integration that is the hallmark of conscious experience. Of course, some of these organisms may *appear* conscious and purposeful. Timelapse videos of slime moulds are particularly hypnotic – though to be fair most things look disconcertingly intelligent in timelapse. Sensory information is distributed in different parts of the organism, and doesn't come together to make the right sort of whole. No whole, no subjectivity.

So we draw a line, below which consciousness is absent. In our judgements of what is and what is not capable of consciousness, we are inclined to lean hard on mechanistic evidence more than behavioural evidence. We are more persuaded by *how* an organism does something than by what it does. We draw our line lower than many do. The neural systems that are capable of forming a phenomenal interface evolved early in the animal line – alongside highly mobile bodies and specialized effective distance senses (Feinberg & Mallatt, 2016; Godfrey-Smith, 2016). They have been highly conserved throughout the vertebrate line and are found in fish, amphibians, reptiles, birds and mammals.

We have argued that many arthropods (including the insects) also have a phenomenal interface, and on that basis we have argued that insects, spiders and crustaceans too are probably conscious of something (Barron & Klein, 2016). Octopus and squid are other likely contenders, but we would need to know more about the cephalopod nervous system to be sure it has a phenomenal interface.

Having a phenomenal interface makes the difference between being aware of things and aware of nothing at all. When we say 'aware', we mean consciousness in its most basic and most primal form. The consciousness of a bee is certainly not as rich and noisily self-reflective as yours. But drawing a line marks an important division: there is something it is like to be a bee – something it is like *for* a bee. There is nothing it is like to be a nematode, a moss, or a rock.

While behaviour alone is not enough to distinguish conscious from non-conscious animals, facts about behaviour and bodies are crucial bits of evidence for what might be conscious. One reason why many authors look to the Cambrian explosion as the moment where consciousness first began is that it is the first appearance of complex animal bodies These bodies had limbs, proper distance senses (such as image forming eyes), and were capable of active movement in all three dimensions. Such bodies enabled predation and other specialised lifestyles.

Having a complex body in a complex environments is what makes the action selection problem hard. Cambrian animals couldn't use the same strategies that *C elegans* uses to solve the action selection problem: now different sources of information need to be weighed off against different needs with complex actions as the solution. Comparing everything to everything gets computationally overwhelming as the types and kinds of information increase. The demand for better, more reliable integration, we suggest, led to the evolution of the phenomenal interface, and hence consciousness.

So when scientists think about whether something is conscious, they need to consider the tripartite links between of what it can sense, what it needs, and how its body allows it to move. The process of consciousness is there to mediate between these three. And that, in turn, gives us a framework to think about even more exotic forms of conscious experience.

♦♦♦

Stanislaw Lem's *Solaris* envisioned a vast planet, covered in an ocean of apparently living protoplasm, shaping and reshaping itself to its own obscure designs (Lem, 1970). Would the planet Solaris be conscious? Lem leaves it deliberately unclear – the perpetually unrealised task of the scientists studying it was to study, as Lem put it, "something that certainly exists, in a mighty manner perhaps, but cannot be reduced to human concepts, ideas or images" (Lem 2002).

It is hard to think about consciousness in the case of something so thoroughly alien because it is completely unclear what Solaris *wants*, or *needs*, or *intends* (if, indeed, any of these things make sense). Similarly, it is hard to know what or how Solaris *acts.* Scientists have thoroughly categorized the intricate surface manifestations of the Solarian ocean; it is the *why* that confuses them.

Confronting the Large Language Models (LLMs) that underpin a lot of current AI can bring on a similar vertigo. Many LLMs *act* intelligent, thoughtful, obsequious, and even conscious. When interacting with them it is often easiest to just treat them as you would another conscious human. Yet we doubt that LLMs are conscious: there is nothing it is like to be one, and insofar as they say otherwise it is a simulacrum of real subjectivity. When you look more closely at what an LLM does, the illusion vanishes.

This is for two interrelated reasons. First, computationally speaking, an LLM does not appear to have the right kind of structure to support a phenomenal interface: it is a completely feed-forward network, optimised for compact representation of very large datasets. The structures that support phenomenal interfaces in biological systems, by contrast, tend to have dense recurrent connections, all the better to bring together multiple streams of information in an ongoing way.

Second, and relatedly, LLMs have no need for a phenomenal interface, because they do not need to sort out the kind of real-time problems of life that a phenomenal interface

supports. They have one job: answering our questions. That job is divorced from the maintenance of whatever counts as their bodies. (Stories where LLMs go rogue to preserve themselves haunt social media. But LLMs are trained on plenty of science fiction; *Neuromancer* cosplay is not self-preservation.) The same questions that are hard to sort about Solaris are easier to approach with LLMs: they have no wants, no needs, nothing that would need the sort of adjudication that consciousness makes possible.

That said, we don't think it's impossible to build a conscious machine. Indeed, we think *trying* to build conscious machines is probably the best thing we could do if we want to really understand consciousness. A phenomenal interface arose as the solution to biological problems. But the right level to understand the phenomenal interface is by thinking about the computations it performs, not what it is made of. This is for good *biological* reasons.

Solving the action selection problem can't be done in a bespoke way for each new organism. Evolution doesn't work like that, and can't work like that – otherwise you'd need to completely rewire the brain each time you added another eye or a new joint or found new types of food. The important thing about a phenomenal interface, as we envision it, is that it *learns* to work with the body and sensorium in which it finds itself. As animals change over evolutionary time – perhaps evolving new senses – the phenomenal interface does not need to fundamentally change. It is a computational structure that can accept a new channel of information without major changes in how it works.

Consciousness is surprisingly flexible even within an individual human life. Work on sensory substitution (Bach-y-Rita, Collins, Saunders, White, & Scadden, 1969; Kaspar, König, Schwandt, & König, 2014) shows that, with appropriate technological scaffolding, people can gain new conscious sensations relatively rapidly. But that is, in many ways, small potatoes compared to growth and development. The journey from infancy to adulthood involves a repeatedly changed body, a storm of new desires as puberty kicks in, and a civilizing process that adds a new layer of needs, desires, and feelings atop it all.

Subjectivity adapts along the way. Something supports that. That something is the same something that supports consciousness in other animals. The best way to characterise that something, we suggest, is not at the level of biological detail but of computational transformations. There is already a lot of variation in the substrate – insects use different neurotransmitters and entirely different neuroanatomy than humans use to solve the action selection problem. The computations of a phenomenal interface could, in theory, end up

being done on a chip. So long as you're able to integrate the right information in the right way, you're conscious, whatever you're made of.

♦♦♦

It's time to pause for some philosophical throat-clearing. One might worry that we have committed the most flat-footed of philosophical sins, implicitly evoking what Daniel Dennett called a 'Cartesian Theatre.' (Dennett, 1991). It's no good to say that integration explains consciousness if 'integration' means something like 'making a tiny model of the world.' Tiny models aren't conscious, so you'd need someone looking at the tiny model. But they would need to be conscious, so would need a tiny model of the tiny model inside the tiny person's head… You see the problem.

The mistake is to think that the process of integration is some kind of precursor, and the *product* of that process is something like a global sensibility or a conscious state or what have you. We prefer views that identify consciousness with the *process of integration* itself: that is, with the process of bringing local sensations into a global context, while simultaneously constituting that global context.

Process metaphysics has a slightly seedy reputation among modern philosophers. Don't worry: there is a sense in which you already believe everything we need. Biology is full of processes, processes all the way down (Dupré, 2013). There are things we naturally call processes: photosynthesis, the Krebs cycle, pregnancy, aging. But even the ordinary objects of biology – the apparently solid furniture like cells, and skin and bone – are processes too. Any bit of your body is constantly being rebuilt and renewed at shorter or longer timescales. Indeed, one of the great achievements of life is to create precariously stable processes out of the constant flux of molecular noise (Hoffmann, 2012).

So think of *being conscious* as like *metabolizing*. There are many substrates and mechanisms necessary for metabolism, and many useful things in turn result from metabolism. But metabolism itself is an ongoing process – an important feature of which is to maintain the conditions under which it can keep happening. There is no further *thing*, The Metabolism, in which metabolism terminates. Nor need one think of a metabolic *state* as anything other than a snapshot of an ongoing process.

Thinking of the action of the phenomenal interface as an ongoing process is not (just) metaphysical fancy footwork. We suggest that the right kind of processes have structural features that are distinctive of consciousness. Consider the idea of a 'point of view,' a key concrete metaphor in thinking about consciousness. Sensory information is processed in the phenomenal interface as having an origin point – a perspectival point. I could even do very simple psychological tests to backtrack the exact point of where I think "I" am precisely relative to things around me, (it's a point in the middle of my head, between and behind my eyes).

Yet none of that requires anything like a *representation* of a point or a map with a 'you are here' pin. Basic transformations – like the ability to correct for self-motion in order to hold a heading – when done consistently, are sufficient to define a single point of view and space around it. Poincaré demonstrated that the algebraic structure determined by the basic operations of self-correction is sufficient to define a geometry with a privileged origin (Poincaré, 1898, 1913). All one needs is knowledge of one's own motion and the ability to separate that from motion of the world. That in turn gives rise to a perceptual distinction between the external world and the subject, and it by necessity gives the subject a perspective on the external world.

We think many of the important structural features of consciousness can be recovered in a similar way. The key to understanding the role of the phenomenal interface is to think of it as supporting a dynamic process, an ongoing moment-by-moment transformation into a common whole. That is why we need a phenomenal interface in the first place, and that process is what it is to be conscious.

♦♦♦

In his 2016 memoir *Being a Beast,* the English author Charles Foster (2016) recounts the heroic and eccentric lengths he went to learn what it was like to be a badger. He dug a burrow in his backyard, ate insects and worms, and emerged only at night to forage. His conclusion was the same: the fact that *he* was doing it was what got in the way.

Badgers are surely conscious. *Whether* something is conscious is a scientifically tractable question. It is often not the question one really cares about. To spiral back around above our starting point, most of our deeper questions come down to *what it is like* to be a bat, an explorer, an addict, or a planet-sized ocean of protoplasm. When one finds oneself

wondering whether something is really conscious, it is often because it very hard to get a handle on what it would be like to be such a thing.

Nagel's challenge is sometimes presented as if it were a fundamental limitation in our powers of unaided imagination. But as Amy Kind points out, imagining is a skill and "Like any skill, one can get better at it with practice, and some people are better at it than others." (Kind, 2020, p141). Take that thought and run with it. Suppose you had any technology you'd need to change your brain activity in whatever ways you'd like (Klein & Barron, 2020). *Then* could you know what it is like to be a bat?

The whiff of paradox returns. I turn on the machine, and have exactly what it what it is like to be a bat. The experience of the bat, in its fullness, is not just what is there but what is *missing*. A bat doesn't know what it's like to be me: to have human senses, human desires, human aspirations to learn about consciousness. The bat-ness must, of its nature, fill me to the brim. The machine is turned off. I return, and now me-ness completely fills the space where bat-ness was. I have not learned what it is like; what space remains for insight to be smuggled back? I *became* a bat, and then again a human, learning nothing along the way.

But of course, this worry quickly generalises. The distance to me in my youth, even to me yesterday, would seem like just as insuperable a gulf. For here an inch is as good as a mile: my experience is always complete. (Or to put it technically – closed under the relationship of co-consciousness (Bayne, 2010): any two experiences that occur together also occur together with any *other* experience that either one occurs with.) If me-now is there imagining me-then, me-now is *there*, in a way that I wasn't then (if it wasn't so, I couldn't wistfully wish that I knew all I know now). So either I've gotten it wrong, or else I've lost myself.

Yet this seems to prove too much. For you do sometimes seem to bridge this gulf. You wake with fragments of dream, in which you were a very different person with very different experiences. Nostalgia can overwhelm you with a kind of double vision of yourself *then* set against who you are *now* (Bailey, 2023). You lose ourselves in a novel and at the end remember where you were even as you return, shakily, to yourself.

What can one learn from such experiences? Start with the human case. Bailey, inspired by Adam Smith, suggests that there is a rich, ethically important phenomenon she calls *humane understanding.* Humane understanding is a type of empathy: it aims in the first place at "the direct apprehension of the intelligibility of others' emotions." (Bailey, 2022,

p 51). Yet, as Bailey rightly notes, we cannot simply do this by trying to feel the same feelings that our target feels. Messner's feeling of triumph and exhaustion atop Everest is reasonable; for me to feel the same thing would not be reasonable, because I am comfortable at the table.

The key, Bailey argues, is to distinguish between *belief-directed* and *thought-directed* emotions: roughly, emotions that are directed at what we actually take to be the case versus emotions that are directed at the things we are thinking about or imagining (2020 4). Empathetic emotion is then "a special category of thought-directed emotion, one that is directed at what we take to be our imaginative recreations of other people's situations" (Bailey, 2022, p 55).

This means that the raw materials for an imagined perspective need not have the same qualities: just as Seurat could use oils to paint the waters of the Seine, so can we use the raw material of our imagination to build worlds that correspond to very different lives. By skilfully transposing from our present circumstance to the imagined one, our imagination furnishes the raw material for this process of understanding. The reactions involved in empathetic understanding, however, are directed at what is imagined. This allows both emotional reaction and distance: one can (must!) have thought-directed emotions without confusing them for ones' belief-directed ones.

From a formal point of view, the process is akin to that by which one computer can emulate a computer with a very different architecture. If you emulate old Atari games on your laptop, your laptop doesn't literally *become* a Atari. Instead, it creates a virtual machine within which the Atari game plays, but over which it maintains full control. That gets the right sort of asymmetry: from the point of view of the virtual machine, the computational world is complete, a Leibnizian monad without windows to the world (Klein, 2020). Rightly so: an Atari doesn't know about laptops and never will. But the laptop has complete insight into the functioning of the Atari game, even though in an important sense it remains wholly different. It creates a whole different world out of its own material.

Simulation gets something of a bad rap in the consciousness literature: paraphrasing Searle, many worry that just as a simulation of a hurricane won't get you wet, a simulation of consciousness wouldn't be conscious (Koch, 2019; Searle, 1980). But the point of empathetic understanding is not, as it were, to evoke or imitate overlapping conscious agents, any more than it is to turn one agent into another. Trying to do either would miss

the point. It is to create, from the point of view of a single conscious agent, the perspective of another conscious agent in its entirety.

The process by which one would go about this sort of humane understanding for other individuals is already familiar. To understand another person we start with the raw material of their life. Return to Keegan's archers at Agincourt. Very roughly, we start at the big picture and work to the particular: the historical circumstance, the life of a medieval with its particular mix of violence and limitation, the campaign itself. If we wanted to imagine a *particular* archer, we'd need more about his life and the details that brought him there. It is that context that gives meaning, and meaning, ultimately, is the foundation of subjectivity.

To go further afield the process is more complex but not different in kind. To understand, say, what it is like to be a bee, you need ethology: what their world is like, what actions are afforded by the environment. You must appreciate basic differences in bodies – so, e.g., the fact that falling is almost inconsequential for insects but the surface tension of liquids is a serious hazard (Haldane, 1926). You need neuroscience. And, as in all science, you probably need mathematics too. The structure of experience has some invariants across species; characterising basic facts at their most abstract level is the realm of the mathematician.

Through humane understanding we can work towards an understanding of others. The greater the difference in experience, the greater the work that needs to be done. Yet the method we use in closer cases can be extended further and further afield. The trick is to remember that every conscious experience comes as part of a detailed whole, and it is the whole that must be constructed to grasp the parts.

♦♦♦

Engineers might build conscious machines someday. Everything we've said above – including scepticism about whether *current* machines are conscious – is entirely consistent with this. Ethicists worry about this possibility, and rightly so. After all, humans regularly make new conscious beings in the old-fashioned way and are not terribly good to one another.

Yet the story we've told also leaves plenty of room for optimism. Humans will have a choice about whether we want to make conscious machines, and a choice about what it

would be like for them if we do. The proposal we have sketched for how to understand consciousness – whether it be that of another human, an animal, a new machine, one's past self – is also one that can be used for design. It is a vision two projects working together as a whole: the scientific and the humane.

On the one hand, the scientific project is to abstract away from particular features to find the general structural principles of conscious experience. An important guide to these principles, in the first instance, will be the computational problems – and particularly those that stem from resolving multiple conflicting bodily imperatives – that the neural bases of consciousness help resolve. As soon as one tells that story about anything other than a developmentally typical adult human, you must abstract to computational principles. Go that far, and you've opened up a lot of scope for what could be conscious.

On the other hand, the *humane* requirement for understanding consciousness is the development of appropriate first-person understanding. It is not just that without trying to know what it's like, one remains sunk in unproductive mystery (though that is also true). If you are dealing with other conscious entities, then you have an ethical obligation to try to meet them on their own ground, and to make their experience intelligible to yourself in as detailed a manner as you can.

The two projects cannot be pursued in isolation from each other. Throughout, we have emphasized the importance of scientific knowledge in fleshing out the imaginative picture. Conversely, the questions raised by humane understanding show gaps in our knowledge: a failure to render someone's sensibility intelligible usually shows that there is more work to be done. We have also gestured at the possibility of more *recherché* fixes: if humane understanding is limited by imagination, we might well look to technological fixes to overcome those limits.

In his reflections on consciousness, Nagel often suggested that the deep problem was not metaphysical but epistemological. Consciousness is about a detailed, particular, meaningful perspective on the world: a view from somewhere. Science abstracts from detail to get you the 'view from nowhere.' Grant this. The gap is still less wide than one might fear. For even if *science* gives an apersonal view, *scientists* are capable of doing two things at once. When you aim to understand consciousness, you don't have to give up yout humanity. Just as conscious experience is, of its nature, an integrated whole, so too does its study need a fully integrated solution.

The drive to humane understanding is also the beginning of the wonder that drives scientific inquiry. Impersonal understanding of consciousness may not be possible. It certainly isn't desirable. To understand consciousness, and to build a future we want to live in, requires developing a rich sensibility that takes sensibility seriously – my sensibility, the sensibility of others, and perhaps the sensibility of the things we have yet to create.

References

Aristotle. (1957). *On the Soul* (W. S. Hett, Trans.). Cambridge: Loeb Classical Library.

Bach-y-Rita, P., Collins, C. C., Saunders, F. A., White, B., & Scadden, L. (1969). Vision substitution by tactile image projection. *Nature, 221*(5184), 963–964.

Bailey, O. (2022). Empathy and the value of humane understanding. *Philosophy and Phenomenological Research, 104*(1), 50–65.

Bailey, O. (2023). What must be lost: on retrospection, authenticity, and some neglected costs of transformation. *Synthese, 201*(6), 189. doi:10.1007/s11229-023-04179-2

Barron, A. B., & Klein, C. (2016). What insects can tell us about the origins of consciousness. *Proceedings of the National Academy of Sciences, 113*(18), 4900-4908. doi:doi:10.1073/pnas.1520084113

Bayne, T. (2010). *The unity of consciousness*. New York: Oxford University Press.

Cobb, N. A. (1914). *Nematodes and their relationships*: US Government Printing Office.

Dennett, D. (1991). *Consciousness Explained*. Boston: Little, Brown, and co.

Dupré, J. (2013). Living Causes. *Aristotelian Society Supplementary Volume, 87*(1), 19-37.

Farrell, B. A. (1950). Experience. *Mind, 59*(April), 170-198.

Feinberg, T. E., & Mallatt, J. M. (2016). *The Ancient Origins of Consciousness: How the Brain Created Experience*. Cambrige Mass.: MIT Press.

Foster, C. (2016). *Being a beast: An intimate and radical look at nature*. London: Profile Books.

Godfrey-Smith, P. (2016). *Other Minds: The Octopus, The Sea, and the Deep Origins of Consciousness*. London: William Collins.

Haldane, J. B. (1926). On being the right size. *Harper's Magazine, 152*, 424–427.

Hamblin, C. L. (1987). *Imperatives*. Oxford: Basil Blackwell.

Hayes, C. F., Rădulescu, R., Bargiacchi, E., Källström, J., Macfarlane, M., Reymond, M., . . . others. (2022). A practical guide to multi-objective reinforcement learning and planning. *Autonomous Agents and Multi-Agent Systems, 36*(1), 26.

Hoffmann, P. M. (2012). *Life's ratchet: how molecular machines extract order from chaos*: Basic Books.

Jackson, F. (1986). What Mary didn't know. *The Journal of Philosophy, 83*(5), 291-295.

Kaspar, K., König, S., Schwandt, J., & König, P. (2014). The experience of new sensorimotor contingencies by sensory augmentation. *Consciousness and Cognition, 28*, 47–63.

Keegan, J. (2011). *The Face of Battle: A study of Agincourt, Waterloo, and The Somme*: Random House

Kind, A. (2020). What imagination teaches. In E. Lambert & J. Schwenkler (Eds.), *Becoming someone new: Essays on transformative experience, choice, and change* (pp. 133–146): Oxford University Press Oxford.

Klein, C. (2015). *What the Body Commands: The Imperative Theory of Pain*. Cambridge, MA: MIT Press.

Klein, C. (2020). Polychronicity and the Process View of Computation. *Philosophy of Science, 87*(5), 1140–1149.
Klein, C., & Barron, A. B. (2020). First-Person Interventions and the Meta-Problem of Consciousness. *Journal of Consciousness Studies, 27*(5-6), 82-90. Retrieved from <Go to ISI>://WOS:000535955900008
Klein, C., & Barron, A. B. (In press). Phenomenal interface theory: a model for basal consciousness. *Philosophical Transactions of the Royal Society B*.
Koch, C. (2019). *The Feeling of Life Itself: Why Consciousness Is Widespread but Can't Be Computed*: MIT Press.
Lem, S. (1970). *Solaris*: Faber & Faber.
Lem, S. (2002) "Solaris by Soderbergh" https://english.lem.pl/arround-lem/adaptations/solaris-soderbergh/147-the-solaris-station (Retrieved 14 June 2025)
Lewis, D. K. (1990). What experience teaches. In W. G. Lycan (Ed.), *Mind and cognition: a reader* (Vol. 13, pp. 29--57): Blackwell.
Merker, B. (2007). Consciousness without a cerebral cortex: A challenge for neuroscience and medicine. *Behavioral and Brain Sciences, 30*, 63-81. doi:10.1017/S0140525X07000891
Nagel, T. (1974). What is it like to be a bat? *The Philosophical Review, 83*(4), 435-450.
Niv, Y., Joel, D., Meilijson, I., & Ruppin, E. (2002). Evolution of reinforcement learning in uncertain environments: a simple explanation for complex foraging behaviors. *Adaptive Behavior, 10*, 5-24.
Paul, L. A. (2016). *Transformative Experience*. Oxford: Oxford University Press.
Poincaré, H. (1898). On the foundations of geometry. *The Monist, 9*(1), 1–43.
Poincaré, H. (1913). *The foundations of science: Science and hypothesis, the value of science, science and method* (G. B. Halsted Ed.): Cambridge University Press.
Roijers, D. M., Vamplew, P., Whiteson, S., & Dazeley, R. (2013). A survey of multi-objective sequential decision-making. *Journal of Artificial Intelligence Research, 48*, 67–113.
Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences, 3*(03), 417–424.
Stoljar, D. (2016). The Semantics of `What it's like'and the Nature of Consciousness. *Mind, 125*(500), 1161–1198.