

The Light & the Room

Invited Chapter: *Introducing Philosophy of Mind, Today*
* Routledge (eds. Devin Curry, Louise Daoust)

ANDREW Y. LEE
University of Toronto, Philosophy

Abstract

To be conscious—according to a common metaphor—is for the “lights to be on inside.” Is this a good metaphor? I argue that the metaphor elicits useful intuitions while staying neutral on controversial philosophical questions. But I also argue that there are two ways of interpreting the metaphor. Is consciousness the inner light itself? Or is consciousness the illuminated room? Call the first sense *subjectivity* (where ‘consciousness’ =_{def} what makes an entity feel some way at all), and the second sense *phenomenal character* (where ‘consciousness’ =_{def} what it feels like to be an entity). I use this distinction—as well as the metaphor of the light and the room—to clarify some philosophical questions about whether consciousness comes in degrees, whether consciousness is multidimensional, and the idea of borderline consciousness.

§1 Introduction¹

To be conscious—according to a common metaphor—is for the “lights to be on inside.” This metaphor is often used to express the sense of ‘consciousness’ that philosophers and scientists are interested in. For an entity to be conscious—in the intended sense—is for that entity to have a *subjective point of view*, meaning that there’s *something it’s like to be* that entity. If so, then the inner light is on; if not, then all is dark inside.

Is this a good metaphor? If you’re like me, you’ll find the metaphor compelling, but you’ll also wonder whether the metaphor might be

¹ This is a final draft of the opening chapter for a forthcoming introductory philosophy of mind coursebook that will be published by Routledge.

misleading in some way. The answer is important. A good metaphor can be a valuable tool for philosophical and scientific insight; a poor metaphor can lead us in the wrong direction and hinder inquiry. Since consciousness is a notoriously puzzling subject-matter—and since the metaphor of the inner light is both popular and intuitive—it's worth thinking about whether the metaphor is useful and illuminating (or confused and obfuscating).

I want to convince you that this *is* a good metaphor. I'll argue that the metaphor elicits useful intuitions while staying neutral on controversial philosophical questions, and I'll argue that the metaphor can sharpen our understanding of what's at stake in some current philosophical debates. If we think of the metaphor as a conceptual tool, then my view is that the tool is both functional and flexible.

Yet I also want to convince you that the metaphor is ambiguous—there are two ways of interpreting it. This may initially seem in tension with my claim that it's a good metaphor. While ambiguity may be a virtue for artistic expression, it's usually a vice for philosophical analysis and scientific inquiry. But I'll argue that the two interpretations correspond to two subtly different senses of 'consciousness', and that identifying the ambiguity enables us to pry apart these two senses. And once we've disentangled the two senses, we'll be able to think more clearly about many theoretical questions about consciousness.

Before addressing the metaphor, though, let me first say a bit more about what I mean by 'consciousness'.

§2 Phenomenal Consciousness

You—at this moment—have a **subjective** point of view. As you read this chapter, it feels a certain way for you to see the page, to hear the ambient sounds around you, to think your thoughts, and to feel your emotions. The totality of these experiences characterizes what it's like to be you right now. This is what philosophers mean by the term **phenomenal consciousness**. More generally, what it is for an entity to be phenomenally conscious is for there to be something it's like to be that entity.

By contrast, there's nothing it's like to be a rock. While the rock has various features—color, shape, density, etc.—it has no subjective point-of-view; it has no feelings of any kind. Most people think that humans, dogs, and octopuses are conscious, that tables, plants, and corporations aren't conscious, and that it's an open question whether slugs, fetuses, and future artificial intelligences might be conscious. Even if you think that some of these intuitions are incorrect, understanding the intuitions helps with identifying the target concept.

The concept of phenomenal consciousness is distinct from the concepts of wakefulness, responsiveness to environment, and self-awareness. If you're dreaming / paralyzed / dazed and confused, then you might not be awake, responsive to your environment, or aware of yourself. But you might still be phenomenally conscious, since there might still be some way it feels for you to be in that state. There are probably important connections between these different senses of 'consciousness'. But it's useful to keep them conceptually distinct, at least at the beginning.

§3 What the Metaphor Leaves Open

Let's return to the metaphor. Occasionally, I encounter someone who balks at the metaphor. But if you're someone who thinks it's misleading in some way, then I invite you to answer: Which aspect of it is misleading?

You might worry that the metaphor invites the idea that consciousness is beyond the scope of scientific investigation. The idea of an inner light may evoke the feeling that consciousness is mysterious or mystical. But the metaphor doesn't say anything about the nature of the light, and it leaves open whether and how the light could be scientifically explained. Think about how physical light itself used to be apparently inexplicable, yet how electromagnetic theory enabled us to understand how light relates to other aspects of physical reality. If you're a **physicalist** then you'll think the light can be explained in physical terms. If you're not a physicalist, then you'll think that the light can't be explained using the kinds of methods and concepts invoked by our physical theories. But either way, you can accept the

metaphor. In other words, the metaphor is neutral on questions about the *nature* of consciousness.

You might worry that the metaphor invites a simplistic picture of knowledge of one's own conscious experiences. The idea of an inner light may seem to suggest that nothing in your own mind is hidden to you (even if it's hidden to others). But the metaphor doesn't say anything about whether or how knowledge of one's own experiences is privileged. And even if knowledge of one's own experiences is special in some respects (perhaps you have direct access only to your own experiences), it might not be special in other respects (perhaps your beliefs about your own experiences are just as fallible as your beliefs about other subject-matters). Whichever way you go, you can accept the metaphor. In other words, the metaphor is neutral on questions about *knowledge* of consciousness.

You might worry that the metaphor invites the idea that the qualities of conscious experiences can exist even in the absence of consciousness. It's natural to imagine the objects in the room still retaining their colors even when the lights are off. But the metaphor doesn't say anything about what happens to the objects when everything goes dark. While you could—as suggested—think of the qualities of the objects as independent of the light, you could also think of the light as what gives color to the objects in the first place. The metaphor enables us to make sense of both pictures without taking a stance on which is correct. In other words, the metaphor is neutral on questions about *unconscious qualities*.

You might worry that the metaphor oversimplifies consciousness. After all, our experiences are rich, complex, and multifaceted. But (as I'll discuss in a moment) the metaphor is neutral on whether the light has different settings: maybe the light can shine more brightly, and maybe the light comes in different colors. And (as I'll discuss in a moment) the metaphor is neutral on what's illuminated: you can think of the ways our experiences feel as corresponding not only to the light, but also to the objects and the space that are illuminated. In other words, the metaphor is neutral on questions about the *structure* and *character* of consciousness.

If the metaphor is neutral on the questions mentioned above, then you might wonder whether the metaphor is doing any work at all. While a metaphor that's too rigid may prematurely close off theoretical questions, a metaphor that's too flexible may lack enough structure to be a useful tool. Over the rest of this chapter, I'll focus on the aspects of the metaphor that I think are illuminating. To start, let me return to the two different interpretations of the metaphor.

§4 The Light and the Room

Suppose we ask: What exactly *is* consciousness, within the metaphor? One answer is that consciousness is the **inner light** itself. Another answer is that consciousness is the whole **illuminated room**. These two interpretations of the metaphor lead to two different senses of 'phenomenal consciousness'.

I'll use *subjectivity* to express the sense of 'consciousness' that corresponds to the inner light. If consciousness = subjectivity, then 'consciousness' can be defined as *what makes an entity feel some way at all*. I'll use *phenomenal character* to express the sense of 'consciousness' that corresponds to the illuminated room. If consciousness = phenomenal character, then 'consciousness' can be defined as *the way it feels to be an entity*.

subjectivity

- *metaphor:* the inner light
- *definition:* what makes an entity feel some way at all

phenomenal character

- *metaphor:* the illuminated room
- *definition:* the way it feels to be an entity

The distinction is subtle. Subjectivity and phenomenal character go hand-in-hand: an entity has what makes it *feel some way at all* if and only if *there's some way that the entity feels*. Therefore, an entity has subjectivity just in case it has phenomenal character. But this doesn't mean that

the properties are identical: two entities could differ in phenomenal character without differing in subjectivity. Even if you and I are exactly the same with respect to the property that makes each of us conscious, what it's like to be you might still differ from what it's like to be me. If these points feel abstract, then think again about the metaphor. The lights are on if and only if the room is illuminated. But even two rooms that are illuminated in exactly the same way could still differ (perhaps in the objects or the space inside).

Subjectivity is itself a dimension of phenomenal character. Because of this, it'd be incorrect to think of subjectivity and phenomenal character as independent parts of consciousness. Instead, the relationship is more analogous to the relationship between color and hue. Everything that's colored has a hue value. And while two objects could differ in color without differing in hue (for example, by differing in saturation or brightness), two objects cannot differ in hue without differing in color. Similarly, every illuminated room has an inner light. And while two rooms could differ without differing how they're illuminated, two rooms cannot differ in their illumination without differing in the way that the rooms are.

If you look at contemporary theories of consciousness, it's usually easy to identify which component of the theory corresponds to subjectivity and which component corresponds to phenomenal character. Let's briefly consider two examples.

1st example: Container Theories: According to some theories, consciousness is kind of like a container. A prominent example is *global workspace theory*, according to which what it is to be conscious is to have a **global workspace** (meaning a central executive system whose information is available for use by a variety of other cognitive systems, such as memory and reasoning). According to these theories, every conscious subject has a container, and what it's like to be a subject is a matter of what's inside their container. Suppose a theory of this kind is correct. Well, what *is* consciousness? If we think of consciousness as subjectivity (the light), then consciousness is the container itself. But if we think of consciousness as phenomenal

character (the room), then consciousness is what's inside the container (alongside the container itself).

2nd example: Awareness Theories: According to some theories, consciousness is a special kind of awareness. A prominent example is *higher-order theory*, according to which what it is to be conscious is to have **higher-order mental states** (meaning mental states that represent other mental states, as opposed to external objects). According to these theories, every conscious subject instantiates the special awareness relation, and what it's like to be a subject is a matter of the objects that one is aware of. Suppose a theory of this kind is correct. Well, what *is* consciousness? If we think of consciousness as subjectivity (the light), then consciousness is the awareness relation itself. But if we think of consciousness as phenomenal character (the room), then consciousness is whatever it is that we're aware of (alongside the awareness relation itself).

If you look through articles on consciousness with a careful eye, you might notice that some authors use 'consciousness' to mean subjectivity while other authors use 'consciousness' to mean phenomenal character. In my opinion, neither of these usages is privileged; both are reasonable ways of understanding the term 'phenomenal consciousness'. In other words, I don't think either subjectivity or phenomenal character has a better claim to being called the *right* definition of 'consciousness'.

So why should we care about the distinction? Well, I care about the distinction because I think certain philosophical questions look very different depending on whether we're talking about subjectivity or about phenomenal character. Let's turn to a few of these questions.

§5 Degrees, Dimensionality, Determinacy

Here are a few big questions about the structure of consciousness (I'll soon explain these structural concepts in more detail):

-
- DEGREES? Does consciousness come in **degrees**?
- DIMENSIONS? Is consciousness **multidimensional**?
- DETERMINACY? Can it be **indeterminate** whether an entity is conscious?

I won't try to answer these questions here. Instead, I want to ask: What's the relationship between these questions? My aim is to argue for two points. First, the questions are logically independent from one another: whether you answer 'yes' (or 'no') to one leaves open whether you answer 'yes' (or 'no') to any other. Second, each question looks different depending on whether we focus on subjectivity or on phenomenal character. I'll make use of the metaphor to argue for both points.

DEGREES?

Does consciousness come in degrees? If 'yes', then some entities are *more conscious* than other entities. Within the metaphor, the relevant question is whether the light sometimes shines more brightly.

Sometimes the question of whether consciousness *comes in degrees* is equated with the question of whether it can *be a matter of degree whether* an entity is conscious. But these are distinct questions. Consider: mass comes in degrees (some things have more mass than others), but it's never a matter of degree whether something has mass (everything either has mass or doesn't). The first question is about whether some entities are conscious to a greater extent than others; the second question is about whether it's always determinate whether an entity is conscious.

You might be tempted to think that consciousness comes in degrees because many features of consciousness—for example, intensity, vivacity, and complexity—come in degrees. But just because some features of consciousness come in degrees doesn't mean that consciousness itself comes in degrees. As a comparison, many features of trees come in degrees—height, age, number of leaves. But treehood itself doesn't come in degrees: one thing can't be **greater in treehood** than another. While it's obvious that

many features of consciousness come in degrees, it's not obvious that consciousness itself comes in degrees.

Questions about degrees of consciousness look different if we're thinking of consciousness as subjectivity vs. as phenomenal character. It's natural to wonder whether the light shines more brightly in some creatures than in others. But it's not obvious what it would mean for the illuminated room to come in degrees. Put another way (using the variables ' x ' and ' y ' to stand for conscious entities), we can make sense of x having more of what makes an entity feel some way at all than y , but it's harder to make sense of *the way in which x is conscious* being greater than *the way in which y is conscious*. This means that when we talk about degrees of consciousness, it's natural to interpret such talk as concerning degrees of subjectivity.

You might be tempted to interpret the size of the room as a measure of degree of phenomenal character. But—setting aside the question of what size of room even represents in the first place—this would yield some incoherent consequences. Remember the heart of our metaphor: to be conscious is for the lights to be on inside. Now imagine a room where the lights are off. If size of room represents degree of consciousness, then that room would represent an entity where there's nothing it's like to be that entity (the lights are off) yet where that entity has a positive degree of consciousness (since any room must have a non-zero size). But what would this even mean? It's reasonable to think that if an entity has a positive degree of consciousness, then there must be something it's like to be that entity.

Some philosophers have argued that consciousness doesn't come in degrees because consciousness is multidimensional. This may sometimes be due to a conflation between phenomenal character (which arguably doesn't come in degrees, and which is clearly multidimensional) and subjectivity (which may come in degrees, and which—I'll argue in a moment—might be unidimensional). But even if we ensure that we're talking about subjectivity (rather than phenomenal character), multidimensionality isn't a good reason for denying degrees. There are many properties—for example, size, health, velocity, and athleticism—that are both multidimensional

and degreed. Imagine, for example, that the light varies in both intensity and aperture. Then it too is *both* multidimensional (there are multiple ways in which the light can vary) *and* degreed (there can be more or less overall illumination).

How would we figure out whether consciousness comes in degrees? Well, I think that what it would *mean* for **x* to be more conscious than *y** would be for **x* to have more of whatever consciousness is than *y**. Since we don't yet know what exactly consciousness is, I don't think we can be sure yet whether some entities are more conscious than others. To find the answer, we first need a better understanding of what the inner light *is*.

DIMENSIONS?

Is consciousness multidimensional? If 'yes', then there are *multiple ways* in which entities can vary with respect to consciousness. Within the metaphor, the relevant question is whether the light can differ along multiple parameters, such as brightness and color.

The term 'dimension' might initially evoke the idea of spatial dimensions. But the intended sense of 'dimension' is more abstract. Think about how color is three-dimensional (the dimensions are hue, saturation, and brightness) even though the dimensions have nothing to do with physical space. A dimension, in the sense relevant here, is a way in which things vary with respect to some kind. Color is three-dimensional because there are three distinct ways in which entities can vary with respect to color. The question, then, is how many distinct ways there are in which entities can vary with respect to consciousness.

Questions about the dimensions of consciousness look different if we're thinking of consciousness as subjectivity vs. as phenomenal character. There are innumerable ways in which phenomenal character varies. Think, for example, about all the different permutations of colors across all the different positions of your **visual field**. Within the metaphor, think about the countless ways in which two rooms can differ. Given this, it's clear that phenomenal character is multidimensional. In fact, it's plausible that

phenomenal character is not only multidimensional, but that it has an extremely high dimensionality.

But what about subjectivity? Well, that's much less obvious. Maybe the light is simply either on or off, maybe it has a dimmer for different brightness values, or maybe it comes in different colors too. These different ways of thinking about the metaphor correspond to different hypotheses about the structure of subjectivity. To figure out what a theory entails about the dimensionality of subjectivity, we need to first figure out what that theory says about which property makes an entity feel some way at all. Then we need to figure out how many distinct ways there are in which that property can vary.

By distinguishing subjectivity from phenomenal character, we can dissolve a puzzle about the question of whether consciousness is multidimensional. On the one hand, dimensions are simply ways in which entities vary with respect to a property, and it seems almost trivially true that there are innumerable ways in which two entities can vary with respect to consciousness. On the other hand, some dimensions of consciousness seem more essential to consciousness than others, and philosophers sometimes entertain the idea that consciousness is low-dimensional.

I think both of these intuitions are on the right track, but they're latching onto different senses of 'consciousness'. The sense of 'consciousness' that's obviously multidimensional is phenomenal character; the sense of 'consciousness' that isn't is subjectivity. While any way in which experiences can vary with respect to what they're like is a candidate for being a dimension of phenomenal character, only a limited class of those ways of varying are candidates for being dimensions of subjectivity. And since subjectivity is what makes an entity feel some way at all, the dimensions of subjectivity are indeed more essential to consciousness. Analogously, there are innumerable ways in which two illuminated rooms can vary, but it's an open question how many ways the inner light can vary. And since every illuminated room must contain an inner light, there's indeed something special about the dimensions of the light.

DETERMINACY?

Can it be indeterminate whether an entity is conscious? If 'yes', then some entities are *borderline conscious*, meaning there's no fact of the matter as to whether those entities are conscious. Within the metaphor, the question is whether the lights are always determinately on or determinately off.

Sometimes the expression 'borderline consciousness' is used as a label for individuals for whom we're not sure whether or not they're phenomenally conscious: for example, coma patients. But this is a different sense of 'borderline consciousness' than the one that concerns determinacy. The question of determinacy isn't merely about whether it's hard to know whether the lights are on or off: even if it's unknowable, there might still be some fact of the matter. And the question isn't merely about whether the light might be very dim: even the dimmest light is determinately on. In other words, even if an entity has only a sliver of feeling, and even if we're not in a position to know that there's something it's like to be that entity, that entity is still determinately conscious.

The existence of borderline cases is often taken to be a symptom of vagueness. Some concepts—such as HEAP, or RICH, or BALD—are *vague* (as opposed to *sharp*). A single grain of sand determinately *isn't* a heap; a pile containing ten thousand grains of sand determinately *is* a heap; but for some cases in the middle, it'll be borderline whether or not a pile containing that many grains of sand is a heap. If you're presented with such a pile and asked whether it's a heap, then it seems inappropriate to simply say 'yes' and inappropriate to simply say 'no'. The reason is that our concept HEAP isn't precise enough to clearly label such cases. If it can be indeterminate whether an entity is conscious, then the concept CONSCIOUS behaves similarly. Maybe our concept CONSCIOUS isn't precise enough for there to always be a fact of the matter as to whether an entity is conscious.

The idea of borderline consciousness may puzzle you. It's hard—maybe impossible—to conceive of borderline consciousness. What would it mean for there to be no fact of the matter as to whether there's something

it's like to be an entity? Yet denying the possibility of borderline consciousness also leads to puzzling consequences. Imagine that we remove one atom at a time from your brain until there's nothing left inside your head. At the start of the procedure, you're determinately conscious; by the end, you're determinately not conscious. But what happens in the middle? If there's no such thing as borderline consciousness, then it follows that the removal of a single atom can be the difference-maker as to whether you're conscious or not. Yet it's very rare to see sharp cutoffs in nature; if you look closely enough, you'll nearly always find shades of gray.

The metaphor of the light tends to generate the intuition that it's always determinate whether an entity is conscious. It's natural to think that the lights are always either on or off, with no borderline states in between. This may initially strike you as a disadvantage: shouldn't the metaphor stay neutral on these sorts of theoretical questions? But I think this aspect of the metaphor is actually a feature, rather than a flaw. As mentioned above, 'borderline consciousness' is sometimes interpreted as expressing something uncontroversial: for example, nobody would dispute that there are patients for whom we're not sure whether or not they're conscious. The metaphor is useful for eliciting the sense of 'borderline consciousness' that's relevant to indeterminacy.

Even though it's intuitive that the lights are always either on or off, the metaphor doesn't rule out the possibility of indeterminacy. Even if it's hard to imagine a room where there's no fact of the matter as to whether the lights are on or not, maybe future insights will help us make sense of the idea. Before modern biology, some had the intuition that it's always determinate whether or not something is alive. But now that we have a better understanding of the nature of life, it's easy to make sense of borderline cases of life (for example, a virus). Although it's hard to conceive of borderline consciousness at present, maybe we'll figure out how to make sense of it in the future. But—at least given our current tools for thinking about consciousness—it's very difficult to understand what borderline consciousness would be.

The question of whether the lights are always either on or off concerns determinacy of subjectivity. But what about determinacy of phenomenal character? You might be tempted to think phenomenal character is indeterminate because some experiences—such as a color experience in peripheral vision, where you see an object merely as colored (but not as any specific shade of color)—represent imprecisely. But just because an experience represents imprecisely doesn't mean that the experience itself is indeterminate. Consider an impressionistic painting: the painting represents imprecisely, but it's still perfectly determinate what the strokes on the canvas are like. Similarly, even if your experience represents imprecisely (say, by representing an object merely as colored), it's still perfectly determinate that you're having an experience with that imprecise content.

You could argue for indeterminacy of phenomenal character by deriving it from indeterminacy of subjectivity. Suppose that it can be indeterminate whether or not an object is illuminated by the light. Then it may be indeterminate what the illuminated room is like *because* it's indeterminate which objects are illuminated. But in this situation, subjectivity is still the fundamental source of indeterminacy. As far as I can see, there isn't a good independent case to be made for indeterminacy of phenomenal character.

Degrees, Dimensions, Determinacy

I've argued that each of our three questions—DEGREES?, DIMENSIONS?, and DETERMINACY?—can be interpreted *either* as a question about subjectivity (the light) or as a question about phenomenal character (the room).

In each case, however, the more interesting question is about subjectivity. It's not clear that it makes sense to think of phenomenal character coming in degrees, it's obvious that phenomenal character is multidimensional, and it's hard to see what would motivate indeterminacy of phenomenal character besides indeterminacy of subjectivity. By contrast, it's an open question whether subjectivity comes in degrees, whether subjectivity is multidimensional, and whether subjectivity can be indeterminate.

I've also argued that these questions are logically independent. Using the metaphor, it's easy to see that (1) whether the light shines more or less brightly is independent from (2) whether the light has multiple parameters, and (3) whether the light is always either determinately on or determinately off. Here are the questions again, paired alongside their interpretation within the metaphor:

DEGREES?

- Does consciousness come in degrees?
 - *Can the light shine more brightly?*

DIMENSIONS?

- Is consciousness multidimensional?
 - *Does the light vary in multiple ways?*

DETERMINACY?

- Can it be indeterminate whether an entity is conscious?
 - *Can it be indeterminate whether the light is on or off?*

Although these questions are logically independent, they might still interact in interesting ways. Maybe you could appeal to multidimensionality and degrees to make a case for indeterminacy. If you accept both determinacy and degrees, then that may constrain what it would mean for something to have a low degree of consciousness. If you reject degrees, then that may constrain what the dimensions of subjectivity might be. If you accept both multidimensionality and degrees, then it might follow that there can be indeterminacy with respect to which of two entities is more conscious. These are all interesting and open questions. But to explore these more complex connections between the questions, we need to first make sure we understand each individual question.

§6 Conclusion

I've defended the metaphor of the light and the room. The metaphor is flexible enough to be neutral on key disputes about consciousness yet functional enough to elicit useful intuitions. The metaphor enables us to disentangle two subtly different interpretations of 'phenomenal consciousness'—subjectivity and phenomenal character. And the metaphor helps us see how that distinction interacts with questions about degrees, dimensions, and determinacy.

There's no substitute for careful philosophical analysis and systematic empirical inquiry. To answer the questions I've canvassed in this chapter, we'll need more than merely a metaphor. But a good metaphor can facilitate analysis and inquiry, and can help us better understand the questions we're asking in the first place. And the light and the room—in my opinion—is a good metaphor.

ANDREW Y. LEE

University of Toronto, Philosophy

Further Readings

Bayne, Tim ; Hohwy, Jakob & Owen, Adrian M. (2016). Are There Levels of Consciousness? *Trends in Cognitive Sciences* 20 (6):405-413.

Churchland, Paul M. (1996). The rediscovery of light. *Journal of Philosophy* 93 (5):211-28.

Kriegel, Uriah (2009). *Subjective Consciousness: A Self-Representational Theory*. Oxford, GB: Oxford University Press UK.

Lee, Andrew Y. (2023). Degrees of Consciousness. *Noûs* 57 (3):553-575.

Nagel, Thomas (1974). What is it like to be a bat? *Philosophical Review* 83 (October):435-50.

Schwitzgebel, E. Borderline consciousness, when it's neither determinately true nor determinately false that experience is present. *Philos Stud* (2023).