



# A Priori Problems with the Metaphysical and Causal Reduction of Consciousness

Kamarajsinh Vaghela\*

## Abstract

Consciousness exists, or so it seems most of the time. However, consciousness is unlike car keys or cellphone in that it is not located at a specific point in space and time. The applicability of physical laws like gravity seems moot at best when it comes to consciousness. What is desirable is an explanation of consciousness that allows it to exist and be part of the very same reality as the car-key or the cellphone, a 'philosophy of immanence' as Gilles Deleuze would put it. The author prefers a view that construes consciousness as *causally-efficacious* (having material effects upon one's body in real-time) and *metaphysically separate* from the brain. In essence, to say that the mind is metaphysically separate from the brain is to deny the proposition that there is nothing more to our subjective experience of mind than the mere activity of the physical brain. This paper looks at a view proposed by John Searle and tries to show that there are empirical problems with a consciousness that is *causally inefficacious* (unable to cause material changes) and *metaphysically identical* (not separate from the brain).

**Keywords:** consciousness, causality, identity, metaphysical reduction, ontological reduction, non-reductive physicalism

## 1. Introduction

Dualism has it that consciousness is part of a reality that is distinct from the 'physical' world that surrounds us. There is a duality of

---

\* Independent Author; [kama\\_bigbang@yahoo.in](mailto:kama_bigbang@yahoo.in)

realities according to this view, one of the “mental” realm and the other of the “physical”. Monism, on the other hand, says that there is just one realm. Two strands emerge out of the monist view - one is physicalism that sides with the scientifically and objectively observable physical world and another is idealism on the side of a purely mental world of subjective consciousness. This essay concerns a middle path, a non-reductive physicalism - one which is physicalist in its outlook - that is it thinks of (experienced) consciousness as no more than the neural activity of the brain- but at the same time also grants consciousness an existence as a thing in itself albeit without causal efficacy - epiphenomenological. The logic behind attributing ‘epiphenomenality’ (or causal inefficacy) to consciousness is that whatever we previously thought that consciousness was causally responsible for can now (given increasing neuro-scientific knowledge) be reduced to the physical activity of the brain. The logic behind attributing ‘existence’ to consciousness is simply that our experience of consciousness is too real to deny it outrightly. A non-reductive physicalist takes consciousness to be real and ontologically irreducible (to be counted amongst all other things that exist in the world), nevertheless inefficacious (without causal function) and metaphysically identical (nothing more than neural-activity).

One non-reductive physicalist is John Searle, who proposes a theory called ‘Biological Naturalism’. It claims that (1) our’s is a physical world that is structured - has lower and higher-level features - consisting of physical particles and forces at the fundamental level (2) and among the higher-level features of the world is consciousness that is causally dependent on the behaviour of lower-level elements. (Searle, 2007)

I argue against such a construal of consciousness because consciousness has causal functions that are not satisfactorily explained by the physical neural processes alone. In this essay, the author develops an analytical critique of the non-reductive physicalism.

Consider a set of H<sub>2</sub>O molecules that constitute the property ‘being liquid’ (in water) or solid (in ice) when organized in some system where, say, the molecules either roll over each other (in case of water) or form lattice structures (in case of ice). According to Searle

(1995), such a property of either 'being liquid' or 'being solid' is caused by and realized in a *causally interacting* system. The molecules interact with each other in a certain way thereby causing the system to behave as liquid or solid.

Now, consider the same set of H<sub>2</sub>O molecules and note that they constitute another property 'being water' without being in any *causally interacting* system. Empirically, 'being water' just *is* 'being H<sub>2</sub>O'. One might say, being ice is also just being H<sub>2</sub>O and the same applies to steam. All are equally true, the distinction between the three forms of liquid, solid and gaseous lie not in their constituent element (same in all three 'H<sub>2</sub>O') but in their "appearance" (as liquid or solid) constituted by the *causally interacting* system (the particular way in which the molecules interact). The properties 'being water' and 'being liquid' are clearly not identical (same as each other): a thing can be H<sub>2</sub>O without being liquid (like ice) and a thing can be liquid without being H<sub>2</sub>O (like mercury). Searle (2007) accepts this non-identity between the properties 'being water' and 'being liquid'.

The distinction between the apparent feature, say, 'being liquid' and the reality 'being water' (which instantiates the apparent feature) is important because it allows us to reduce the apparent feature to the lower level without having any implications on reality. We can naturally reduce the property of being 'liquid' to the causally interacting microphysical system without reducing the property of being 'water'. In the case of consciousness, however, its appearance *is* its reality (Searle, 2007). This nuanced understanding of water that distinguishes its appearance from its reality allows us to conceptualize the metaphysics of matters other than water, say, consciousness. However, consciousness is not like water, in fact, it is unlike most things that exist in our world. In the case of consciousness, its very appearance *is* its reality. We know of consciousness only through its appearance to us - there are no objective tools or methods to bring consciousness to the fore, say, on a petridish under a microscope. The very essence of consciousness is realized in its appearance to the subject's mind.

Searle has three major commitments in regard to consciousness: 'causal emergence', 'causal reduction' and 'ontological irreducibility'.

## A.Causal Emergence

The reducibility of the property 'being liquid' to a *causally interacting* system makes it a "causally-emergent" property. Searle calls it causally emergent because the property that is reducible to the interacting system is *not* reducible to any of its individual constituent parts. The property 'being liquid' is not reducible to a single molecule but only to a set of molecules in a certain interaction. Analogously, for Searle (1992), consciousness is reducible to the neurophysical processes but not to any individual neuron; it is causally-emergent.

## B.Causal Reducibility

The property 'being liquid' seems to have causal properties that 'being solid' does not have; for example, 'liquid' can *flow*. But all causal powers such as the property of flowing ascribed to 'being liquid' are explainable in terms of the underlying microphysical system. In other words, 'being liquid' has no novel causal properties other than those possessed by its underlying microphysical system. For Searle (1992), consciousness has no novel causal powers; it is causally reducible.

## C.Ontological Irreducibility (standard version - not Searle's)

There are properties that can be ascribed to 'water in a liquid state' that cannot be ascribed to either individual microphysical elements of water or to the causally interacting system *per se*. 'Water in a liquid state' has the causal property of quenching my thirst, while neither an H<sub>2</sub>O molecule by itself nor the causally interacting system of something other than H<sub>2</sub>O has the same causal property. The quenching of my thirst is an *emergent* property on the property 'being water in liquid state'. A property is genuinely ontologically irreducible (or ontologically emergent) only if it is also causally irreducible. Kim's Alexander's dictum says "for something to be real, it needs to have causal powers" (Kim, 1992, p. 134). If it is true that the mark of an ontologically real thing is that it has novel causal powers, then for consciousness to be ontologically real it must be causally irreducible (have novel causal powers).

#### **4.Ontological Irreducibility (Searle's 'metaphysically trivial' version)**

Searle subscribes to what is called 'weak emergence'. Epistemological or weak emergence claims that the higher level of reality (like consciousness) is just a level of description, metaphysically and causally identical to the lower level (no more than neural activity in the case of consciousness). Admitting things like consciousness into some ontology (a list of things that definitively exist in the world) has no metaphysical implications. Consciousness, for Searle (2007), is irreducible for "trivial reasons." By 'trivial' he means that consciousness, since it cannot have causal efficacy in the material world, has no meaningful grounds on which it can claim metaphysical separateness from the brain.

I'll now state what I think is obvious (perhaps even redundant) as my capsule-version argument, unpacking it in the rest of the essay:

- I. an ontologically real object, like water or wood, can either have casual properties identical to the causally interacting microphysical level or have novel causal properties (not reducible to the micophysical level). Having novel causal properties amounts to genuine ontological irreducibility - like we saw in the case of 'water in a liquid state' quenching my thirst (going by Kim's 'Alexander's Dictum'). Having identical causal properties to the lower level amounts to causal reduction - the property of 'flowing' in liquid can be easily reduced to the causally interacting system. If flowing water causes a thing to float we can be sure that the molecular interaction between the flowing thing and floating thing make the floating possible. It is clear that this is an either/or case between being causally irreducible (thus ontological irreducible) and causally reducible, but Searle tries to have both at once.
- II. For an apparent feature, like the property of being 'liquid' or 'solid', having novel causal properties is to be causally irreducible. However, having identical causal properties to the lower level amounts to both causal and ontological reduction to the lower level (the apparent property of 'flowing' and all its causal efficacy is no more than the

causally interacting system). This too is clearly an either/or case, meaning if 'not causally irreducible' (not having novel causal powers) then 'ontological reduction' (not a thing in itself) follows.

This is why I think causal reducibility should be given up in the case of consciousness, given that its appearance *is* its reality. I agree with Kim when he says, "in an attempt to explain mental causation, [reductionism] all but banishes the very mentality it was out to save." (Kim, 1995, p. 194).

The growing influence of physical sciences (within and outside of philosophy) vigilantly defends against any possibility of what is called 'overdetermination'. Overdetermination simply is the case of one effect with more than one immediate cause: so if a ball is thrown in the air with some force, its only cause must be (according to physical sciences) the force of the thrower's hand and its physical movements. Trying to attach another force as the immediate cause of the event leads to overdetermination - a (physical) impossibility. Another concept is equally well guarded by the influence of physical sciences, that of 'physical closure'. The concept of 'physical closure' maintains that a physical effect must have a physical cause. Any non-physical cause (like consciousness) is not permitted to enter the loop of causality.

Now since Searle rhetorically asks and answers, "But are you saying that the conscious state itself has electrical charges and has spatial dimensions in the brain?... obviously yes" (Searle, 2007, p. 176) there are two questions worth asking. (1) wouldn't causal efficacy of the mental (if it is indeed a part of the causal chain and so ontologically irreducible) require that it has some property that is causally relevant (does the causal work) without invoking overdetermination? (2) Is it possible for something non-material to possess material properties, or for a non-material property to be causally relevant without violating physical closure?

Searle answers the first question by positing causal reducibility of consciousness to the neurophysical brain.

"..the reason consciousness can function causally in a "physical" world is that it has physical properties. Every conscious state is realized in a certain physical structure in the brain and has the

conscious powers of all of these physical structures.” (Searle, 2007, p. 176)

To answer the second question, about causal relevance without violating ‘closure’, Searle takes a longer route. He makes a list of material things like ontological objectivity, quantitative measurability, spatial location and so on. He then contrasts this list with the features of consciousness, like ontological subjectivity, qualitativensness, intentionality, and others. Such a contrast makes it clear that in no way can consciousness, without spatiality or any physical features, function causally in a physical world. But evidently, consciousness does have causal efficacy, in the form of what Searle calls intention-in-action, and to bridge this intuitive gap Searle gets rid of the contrast itself. Once the opposing terminologies “material” and “mental” are done away with, we have exactly one ontological domain that consists of both, the features of consciousness and their (neuro)physical counterparts. Searle suggests that all the non-material features of consciousness, rightly classified, are ‘physical’, and so they may be causally efficacious,“ without violating the causal closure, overdetermination, epiphenomenalism or any other paradoxes.” (Searle, 1995, p. 219)

If these are Searle’s answers to (1) and (2), then the circular reasoning in them must be pointed out which render his arguments futile; (according to Searle) consciousness can be causally efficacious because it has physical properties and since consciousness can be classified as physical, there is no problem with it having causally relevant properties. If this reasoning is granted to Searle then the following can be inferred from his “causal reducibility” proposition: consciousness is *causally efficacious* so far as it is physical, and since all the causal work attributed to consciousness is sufficiently explained by the workings of the brain, consciousness is *causally reducible* in so far as it is metaphysically identical to the brain. Overlooking this circular reasoning in Searle is an option that should not be available to us or even Searle himself. This is another clue that the pursuit of drawing up a consciousness without causal efficacy should be given up.

This section argues in favour of the non-reducibility of causal powers. Most philosophers recognize that “in the mind-body

problem three players are on the scene, not two; they are [mental states], the brain, and behaviour.” (Kim, 2006, p. 195)

According to Searle, consciousness has intentionality which is intrinsically connected to action, “Intentionality is directedness and intending to do something is just one kind of Intentionality among others.” (Searle, 1980, pg. 3) ‘Intentional content’ is the intentionality directed towards the objects in the world and ‘intentional components’ are feelings like desire and belief with which intentional content is presented. Searle explains ‘intentional components’ by reposing Wittgenstein’s question: If I *raise my arm*, what is left over if I subtract the fact that my arm went up?” What is left over is the ‘intentional component’ which involves “certain feelings and bodily sensations.”

‘Intentionality’ has phenomenal and logical properties; a phenomenal property is simply the *feeling* of believing, desiring or intending, and a logical property is the *condition of satisfaction* for that belief to come true, desire to come to pass or intention to have a corresponding action performed (Searle, 1980). My ‘seeing a red apple’, for example, has the phenomenal property of ‘redness’, ‘roundness’, etc., and the logical property of *being caused by* (there being) a red round apple present in my field of view. The movement to perform some action, in order for that movement to count as action per se, must satisfy the condition of self-referentiality, the logical property of being (movement, in this case) caused by there being the intention to perform *that* very action.

Searle suggests that a phenomenal property (say, the *belief* that I) with the logical property (say, am *thirsty*) is ‘caused by and realized in’ the brain, and so causally reducible to the neurophysical processes. The problem, as I see it, is that neurophysical processes alone cannot explain action if a movement to genuinely count as action must satisfy a logical self-referentiality condition of ‘*being caused by there being*’.

If I am feeling thirsty and need to drink water, for my movement (towards the kitchen or wherever the water is) to count as action, that movement must be caused by the very intention of drinking water. Movement, however, consists of several constituent intentions like intention to ‘move in a certain direction’, ‘at a certain



speed', 'in a certain way' and so on. These constituent movements do not necessarily satisfy the logical condition of being directly caused by the intention of drinking water, since they are subject to change or modification. Upon feeling thirsty I may start walking towards the kitchen thinking that the water is in the kitchen, but will change the direction of my movement if on my way I am reminded of the fact that the water is in the bedroom. This change in movement does not mean a change in action or a change in intention. The 'being caused by being there' condition of this change in movement is not water or even the feeling of thirst, instead, it is simply the knowledge of the location of water.

Fred Dretske's passage reiterates my point: "After all, the dog is here and the tree is there; surely the cat's knowledge of this fact is relevant to why it runs in a certain direction. Its genes and hormones won't help us explain this aspect of the cat's flight however much they help us to understand why cats are afraid of dogs or why they tend to run in such situations." (Dretske, 1988, p. 32)

To explicate the point further, take the same cat and put her in a more testing (similar but distinct) situation, say, it realizes midway that the thing it was running towards (thinking it is a tree), is just another weird-looking dog. That cat's coming to knowledge about this new information must surely be relevant to her revised direction of flight. In both Dretske's example and my version of it, the cat's phenomenal property must be the same viz. 'fear', and also the logical property corresponding to that fear must be the same viz. 'dogs'. However the physical movement in both examples will be different because of its constituent parts. Neurophysical processes may account for movement as action against the backdrop of a specific intention, but that is only too rigid a formulation of movement because it has constituent parts that are subject to sudden and almost intuitive change and modification.

By a causal reduction to neurophysical processes we are restricted to an explanation of materially physical movements alone and lose out an explanation of non-material aspects of action.

Searle concludes his essay 'Dualism Revisited' with the following argument:

1. Materialism says truly that there is only a physical world made of physical particles in fields of force and says falsely that consciousness, as irreducible, qualitative subjectivity, does not exist.
2. Dualism says truly that consciousness is irreducible as qualitative subjectivity i.e. cannot be reduced to a third-person phenomenon and says falsely that irreducible consciousness is something over and above the physical world.
3. Thus, taking what is said truly and leaving what is said falsely, consciousness exists as a real, irreducible phenomenon [as subjectivity] in the physical world but not as something over and above the physical world.

Given that the appearance of consciousness *is* its reality, the appearance is deemed irreducible because its reduction implies a reduction of its reality. If so, then two questions naturally arise: "what property does its reality constitute of in virtue of which it *cannot* be reduced?" and, "what property does the appearance constitute of in virtue of which it *should not* be reduced?" The answer to both these questions, since appearance *is* reality in case of consciousness, must be the same.

If Searle wants to claim that consciousness is ontologically irreducible (a thing in itself) while also being metaphysically identical to (nothing over and above) the brain processes he must give us at least one property that can be ascribed to consciousness but not to the brain processes. That property, according to Searle, is subjectivity. Naturally, subjectivity or first-person-ness is a property attributable to consciousness but not its underlying neurophysical processes. However, if the appearance of consciousness truly just *is* the reality of consciousness then Searle must also point out a property common to both in virtue of which they must not be ontologically reduced. Searle fails at this important task of pointing out a common property: to the appearance of consciousness Searle says it has subjectivity and to

the reality of consciousness Searle says it has physical dimensions in the brain.

If this is the case then Searle's view admits to the proposition that the appearance of consciousness is **not** the very same thing as its reality. Searle, I assume, will not want to take this lane which goes down towards dualism, a disaster at the least.

## References

- Dretske, Fred. *Explaining behaviour: Reasons in a world of causes*. The MIT Press, 1988.
- Kim, Jaegwon. Being realistic about emergence. In P. Clayton & P. Davis (Eds.), *Re-emergence of emergence: The emergentist hypothesis from science to religion*. New York: Oxford University Press, 2006. 189-202.
- Kim, Jaegwon. Downward causation in emergence and non-reductive physicalism. In A. Beckermann, H. Flohr, J. Kim (Eds.), *Emergence or reduction? Essays on the prospect of nonreductive physicalism*. Berlin, New York: de Gruyter, 1992. 119-138.
- Kim, Jaegwon. Mental causation in Searle's 'Biological Naturalism.' *Philosophy and Phenomenological Research*, Vol. 55, no. 1, 1995, 189-194.
- Searle, John. Consciousness, the brain and the connection principle : A reply. *Philosophy and Phenomenological Research*, 55, (1995), 217-232.
- Searle, John. Dualism revisited. *Journal of Physiology Paris*, 101, (2007), 169-178.
- Searle, John. The intentionality of intention and action. *Cognitive Science*, 4, (1980), 47-70.
- Searle, John. *The rediscovery of the mind*. United States: The MIT Press, 1992.