

The Travails of the HOT Theory of Consciousness

Dr. Aiman Reyaz

Assistant Professor of English

Ram Jaipal College

Jai Prakash University, Chapra, India

aimanreyaziitp@gmail.com

&

Mr. Samar Reyaz

Assistant Professor of Management

St. Xavier's College of Management & Technology, Patna, India

sam.reyaz@gmail.com

Abstract

It is a matter of extreme inquiry as to how a thing which is purely physical can conceive and develop conscious experiences. There appears to have a void in the marriage between physical (brain) and the mental (mind). Many theories try to bridge the vacuity between the two but two contemporary theories have emerged quite stronger and saner than others: a. the higher-order thought (HOT) theory and b. the anti-theories. The purpose of this paper is to trace the various evolutionary routes the theories of consciousness have taken and to elaborate mostly on the above mentioned two theories. The researchers have used descriptive and deductive reasoning to argue the points.

Key Terms: Physical/Mental; Consciousness; Theory; Reasoning

HOT Consciousness

The hard problem of consciousness is a very relevant issue. How can any physical phenomenon in the brain explain conscious experience, as we know it? What would a physical understanding of subjective quality even look like? Between the physical and the mental, there still seems to be an explanatory gap. Francis Crick's (1916-2004) and Roger Penrose's (b. 1931) attempt to supply a physical explanation of consciousness, but neither of which seemed to close the explanatory gap (Penrose 16). David Chalmers (b. 1966) also tried to resolve the issue by

making a radical proposal that consciousness should be recognised as a fundamental principle in the universe (Chalmers 114). But even that does not really give us an explanation. In this paper, the researchers want to consider two further approaches to the problem; one of these is a very different kind of theory – the higher-order thought or HOT theory of consciousness. It is perhaps the strongest contender for a fully functionalist theory of consciousness (Lacewing 39). The other approach is represented not by a theory of consciousness but by a handful of Anti-theories. There the thesis is that consciousness will forever be beyond scientific explanation; that an understanding of consciousness will forever escape our grasp.

First, the HOT theory: The materialists or functionalists response to the hard problem of consciousness is two-fold; one was hope for a miracle and another was a deflationary response. Regarding the first issue, while exploring the neuroscience of the brain we come across a physical structure or a physical process that not only correlates with consciousness but that suddenly makes consciousness understandable for the first time. That is where both Crick's and Penrose's theories belong. They attempt at showing some physical miracle that explains consciousness and unfortunately neither seems to really give us the miracle we need. The other *Deflationist response* is that consciousness is not some strange thing, not some hidden brain process, but a particular concept (Hale 231). But we need to understand first and foremost our concept of consciousness. Once we really understand the concept, we will see that the hard problem is not really so hard after all. That is what David Rosenthal's Higher Order Thought (HOT) Theory attempts. The HOT theory aims at a more complete conceptual understanding of consciousness in terms of the functional organisation of mental states. Rosenthal's theory is an attempt to better understand what exactly our concept of consciousness is (Rosenthal 20). It starts with some conceptual distinctions intended to help us put our finger on precisely the kind of consciousness we are talking about.

Here is a first distinction, sometimes we talk about a person or an animal being conscious as opposed to being comatose or asleep. What do we mean when we say that the patient is conscious? We just mean that he is awake and responding. That looks like precisely the kind of thing for which we could offer a functionalist account. We will call this concept- creature consciousness. But sometimes we talk about something different, we talk about not an animal or a person but a mental state as being conscious. I am sure this has happened to you. You pull into

the parking lot at your destination and realise that you have no memory of the drive. That does not mean that you were in a comma. You could not have gotten here if it were. But your driving was in some sense unconscious; you were driving on automatic pilot. The mental states involved in your driving weren't conscious mental states. That second sense of the term we will call state consciousness. State consciousness is what the hard problem is all about. That consciousness that you have now but you lack when you are driving zombie-like in automatic pilot, that is what the hard problem is all about and that is what Rosenthal's theory is designed to explain.

Distinctions

The first distinction is between creature consciousness and state consciousness. There is a second distinction that will be important. Sometimes we speak of being conscious *of* something; of being conscious of what someone is saying, for instance; of being conscious of how many people are in room; or even conscious of the time. That kind of consciousness *of* is what Rosenthal calls, transitive consciousness. It is like a transitive verb; Joe hit Mary. Transitive consciousness has you on one side and the thing that you are conscious *of*, on the other. With an eye to the hard problem, it is not merely consciousness *of* that we are out to explain but what we are interested in is whether the mental state itself is conscious (Mandik 30). We speak for example of a belief being a conscious belief, of your anger being conscious anger, and of your awareness being conscious awareness. That is what Rosenthal calls intransitive consciousness – the conscious character of the mental state itself. With these two distinctions, we have identified our target. And in trying to answer the hard problem of consciousness, what we want to understand is not creature consciousness but state consciousness. What we want to understand is not transitive consciousness but intransitive consciousness, the fact that the mental state itself is a conscious mental state. What we want to understand is the subjective experience of intransitive state consciousness.

The HOT theory is a deflationary theory. Here is the balloon, the false theory that Rosenthal's sets out to deflate. When it is asked, what is the difference between a conscious mental state and an unconscious one; we tend to think that the conscious mental state has some special intrinsic quality, it somehow glows. When one is driving on an automatic pilot, one evidently sees the kid on the tricycle, after all one slowed down. But the person was on automatic pilot; the perception was not conscious perception. Right now, for example when one

asks you to pay attention to the text in front of you, you can see it consciously. What is the difference between those two cases? The intrinsic quality theory which we tell ourselves is this, when we consciously see something, it glows for us. The lights are on; the mental state has a different property: a particular glowing quality, the quality of consciousness. That is what makes the hard problem seem hard. Where does that extra quality come from? We focus on that mysterious quality and we try to explain the mysterious glow. But we are fooling ourselves. If we want to understand intransitive state consciousness and we do, we should stop thinking in terms of some kind of intrinsic property, some kind of glow. That is not in fact what it is for a mental state to be a conscious mental state. Here is the higher order thought theory- for a state to be a conscious state is not an intrinsic property, it is a relational property, a relation to something else. What it is for a mental state to be conscious is for there to be some other mental state, a higher order thought that is about it. To be conscious is not something a state has, it is something it gets. A mental state is conscious just in case it is the target of some other mental state.

What makes a state conscious is not some kind of glow, but what makes it state conscious is simply that it is on the receiving end of some other state, a higher order state. It is when you are attending to your own mental states that they are conscious, it is when you are contemplating your own anger, it is conscious anger; it is when you realise that you have a particular belief, it is a conscious belief; it is when you are thinking about perceptions, then they are conscious perceptions. That is all of subjective experience is. To say that a mental state 'A' is conscious is just to say that some other mental state B is piggy backing on it. That there is some other mental state that is targeting it, that higher order state might itself be conscious. There might be some still higher order states that targets it. But it does not have to be. There are some additional bells and whistles but that is the core of the HOT Theory. The idea is that we are conscious *of* lots of things in a purely functional way: conscious of the cars swirling in front of us, conscious of the kid on the tricycle. That is just transitive consciousness; it is a functional directedness towards environment. That is something you have even when you are driving on automatic pilot. But sometimes what we are conscious *of* is another mental state, our anger or a particular belief or our current perceptions. When that happens, intransitive consciousness happens. That is when our mental states are conscious, when we are conscious *of* them. Not because they glow on a particular way but simply because they are something we are aware of. That is the reason why it is called a deflationist or deflationary theory; it tries to deflate the concept of consciousness

(Bermudez 9). Consciousness is not some kind of intrinsic glow a state has. When we are talking about consciousness we talk about a relational pattern between mental states; the fact that one mental state, the conscious state is the target of another. It is just a relational pattern between mental states.

Consciously Conscious

This also ends up being a purely functionalist theory. Whatever the mechanism of thinking about something might be, consciousness is analysed as what happens to a state when it becomes that something you are thinking about. Consciousness is merely a matter of the functional organizational mental states, after all. There is an interesting tying here to a different theory. Harry Frankfurt offers a hierarchical theory of free will. To be free is not merely to act in accordance with your desires; it also demands that you have the desires that you desire (Costa 176). Freedom is a matter of higher order desires- desires about desires. In the HOT Theory, consciousness is a matter of hierarchy as well. Consciousness is a matter of higher order mental states- mental states about mental states. Free will and consciousness do seem to have something in common. It is perhaps a point of favour of both Frankfurt and Rosenthal's theories that they seem to operate in parallel. How good is the theory? One has to admit that one gets in two minds about it. On the positive side, this is the closest we have to a successful deflationary theory of consciousness. It is the closest functionalism has come to offering a genuine analysis of the concept of consciousness to a genuine conceptual understanding of consciousness. Many things about the theory seem right. Consciousness, in precisely the sense that we are after, does come with an awareness of awareness, just as the theory says.

The relational characteristics that Rosenthal points to between the mental states, seem to be a feature of subjective experience. On the negative side, there may be a problem with the theory; it may be sneaking in what it is trying to explain through the back door. Here is a potential problem: it turns out that in order for the theory to work it has to exclude certain cases. There are cases for example, in which I have a belief about my own beliefs but in which those beliefs don't thereby become conscious. There are cases in which I have thoughts about my own perceptions but in which those perceptions don't thereby become conscious. Suppose I am not consciously angry but you can tell from the way I grit my teeth and clench my hands that I am angry and you tell me so. At that point, I know I am angry. Does it follow that my anger is

suddenly conscious, I can suddenly feel it. Sometimes the answer is 'yes'; you tell me I am angry, I suddenly realise my anger and I am flooded by the feeling of anger. But sometimes the answer is 'no'. You tell me I am angry, I am now aware of my anger, I can see it in my own behaviour but I still don't feel it. Perhaps I know I have reason to be angry. I may know that I am acting angrily but somehow I just don't feel angry. That would be a counter example to the theory.

The HOT Theory says that the mental state is a conscious mental state, that anger is conscious anger whenever there is a higher order mental state about it. But in this case, the person has a higher order mental state about my anger. One realises one is angry but one's anger still is not felt as conscious anger. In order to avoid that kind of counter example, the theory will have to be refined. The theory will have to say that a mental state is conscious not when there is just any mental state about it, in just any old way but when a certain kind of higher mental state is about it, in a certain way. Rosenthal sees the problem and he offers a refinement of the theory. His refinement is to say that the higher order thought cannot come from external observation by observing my own behaviour, for example. The higher order thought cannot come from inference or from you telling me. The higher order thought has to attach to the other one in some immediate way. The full theory then has to have an immediacy clause. A mental state is conscious when there is a higher-order mental state about it in some immediate way. The question that arises is what kind of connection counts as immediate? Does that immediate really mean that the lower state has to be directly perceived, has to be directly felt, has to be directly sensed. That makes it sound like the lower state would have to have a particular felt quality, it would have to glow. That glow is precisely what the higher order thought theory was supposed to get us away from.

So, the worry is that the theory may merely be smuggling the glow of consciousness in the back door by way of the immediacy clause. If so, of course the theory is not explaining what it is supposed to explain and is not deflating what it is supposed to deflate. The authors would like to re-emphasise the positive; the HOT theory is the closest we have to a genuinely Functionalist theory of consciousness. Just as the empirical theories on the table deserve further empirical investigation, this distinctly conceptual theory deserves further conceptual exploration. We have looked at a range of theories of consciousness but there are also important anti-theories.

They don't offer us a solution to the question of consciousness. They don't even suggest where the answer may lie. But these anti-theories tell us that there is no answer to the question of consciousness or at least no answer that we could ever possibly find. The punch line of the anti-theories is that consciousness must forever and inevitably remain a mystery (Hartman and O'Hara 180). Those who propose these theories are sometimes called 'mysterions' for precisely that reason. The 'mysterions' anti-theories come in different strengths and even those who propose them offer a different strength in different times. The paper is going to outline two such theories, related to but in some ways distinct from each other. The two anti-theories are of Thomas Nagel and Collen Magyen.

The Anti-Theories

Nagel is the man who asked what it is like to be a bat, a very effective way of pressing the hard problem of consciousness (Murtagh and Lee 83). But Nagel did not merely ask the question, he used it as a way of offering a particular view of the nature of the problem at issue, an anti-theory of the problem of consciousness. Here goes, there are two different perspectives that we take toward the universe: one perspective is essentially Subjective, which simply refers to my perspective from my experience, my sensations and my point of view. That is the subjective perspective on the universe. The person can share how he/she felt when he/she got the wonderful news or how painful the person felt he/she remembers a particular time in the past. The other perspective is Objective, which does not characterize itself in terms of any particular point of view (Cavanna and Nani 61). When it is stated that the moon orbits the earth or that the Sun is 93 million miles away from Earth or that granite is an igneous rock, it does not mean that one is reporting anything personal or subjective experience. Rather it denotes the bare objective facts, the way things *really* are, not merely the way someone perceives them to be. This is the objective view intended to transcend our individual perspectives.

Objectivity is what Nagel calls *The View From Nowhere*. Nagel thinks that objectivity is a major conceptual achievement, essential to the notion of science. He thinks that we should strive for a wider and more inclusive objectivity in a number of areas (Nagel 68). But he does not think that objectivity is the whole story. We make contact with the universe through both, an objective conception and our individual subjective points of view. Many aspects of the universe may be accessible through both roots. One can glimpse the moon from a particular point of view

at a particular time and still understand it as an objective phenomenon independent of me. But some aspects of the universe will not be accessible through both roots. Some aspects of the universe may only be accessible objectively, some aspects of the universe may only be accessible subjectively and one of them is consciousness. Consciousness is a genuine and fully real phenomenon in the universe but science cannot seem to capture it. This is because it is an essentially subjective aspect of the universe. Something essentially subjective, something accessible in principle only from the subjective point of view is something one could never see in *The View From Nowhere*. It is something one could never capture in a perspective that wasn't from a particular point of view. Nagel is forthright in drawing the logical conclusions of such a view. He is a committed realist. There is indeed reality out there independent of our involvement and capable of being described and understood objectively. But Nagel says there is more to reality than objective reality. Reality has both an objective and a subjective side; and science, precisely because it is objective to the core can only give us access to half. Why haven't we managed to get a scientific explanation of consciousness? This is because a scientific explanation would be an objective explanation, something free from any particular point of view. But what we are trying to explain is essentially subjective, something that is only visible from a particular point of view.

No wonder no scientific explanation has been forthcoming. Nagel holds a similar view regarding the issue of free will; free will is something conceived from a subjective point of view, from the point of view of someone having to make a choice. That is why it seems to clash with an objective view of the universe in terms of physical laws, and causes and effects. As has been argued that the anti-theories come in a number of different strengths, so how strong is Nagel's claim? Might our progress toward objectivity eventually conclude the subjective in a way that would make a theory of consciousness possible? Nagel offers different prognoses to different times. At some points, he is optimistic about the possibility of an eventual theory. He says:

“What is needed is something we do not have: a theory of conscious organisms as physical systems...which also have an individual perspective on the world... the strange truth seems to be that certain complex, biologically generated physical systems, of which each of us is an example, have rich nonphysical properties. An integrated theory of reality must account for this, and I believe that if and when it arrives, probably not for centuries,

it will alter our conception of the universe as radically as anything has to date.” (Nagel 51)

Nagel uses an analogy; the physics formulated by Sir Isaac Newton is what we think of as classical mechanics. It offered a new understanding of physical phenomenon from the behaviour of gases under pressure, to the trajectory of a bullet, to the orbit of the planets. In time, it became clear that some things did not quite fit. People knew about magnetism, electricity but neither of them had a ready explanation within classical mechanics. James Clerk Maxwell developed a concept of the electromagnetic field, but he thought it could eventually be explained in classical mechanics. The work of Hendrik Lorentz made people realise that it would not be; that electromagnetism represented a fundamentally distinct aspect of reality and we would have to change our physical theories in order to incorporate it. Only with Albert Einstein did we get a new physics that made it clear how the range of physical forces fit together as a whole. Could the same thing happen with consciousness? Could we be now where a classical mechanics was at the time of Newton, aware of the phenomenon of consciousness but without a Maxwell and a Lorentz to give us the concepts we were willing to handle it, waiting for an Einstein to show us how the physical and the mental are really part of the same whole. That is the prospect that Nagel holds out as optimistic mode but one has to admit that it is just another form of hope for a miracle, a miracle of radical conceptual change that will suddenly let us see how the physical and the mental fit together.

In his less optimistic moments Nagel suggests that the two perspectives may be forever irreconcilable. How could one get anything like an objective grasp of an essentially subjective fact? In those moments Nagel suggests that our failure to understand consciousness is not merely an unfortunate aspect of our point of time, something we don't yet know, an aspect of our historical limitations. In those moments he suggests that consciousness will be something it will forever be impossible to understand. Our inability to understand consciousness is not merely a passing aspect of our historical predicament but a permanent aspect of the human predicament. That pessimistic proposal is carried further in the work of Colin McGinn. Can we solve the mind-body problem, he asks? His answer is No. He thinks there must be a physical truth about consciousness, but it is a physical truth that it will be impossible for us grasp. The first thing McGinn wants to convince you of is something weaker. Isn't it at least possible that the answer

to the hard problem will forever be beyond human reach? Human beings approach the universe in terms of human concepts, in terms of concepts that we have developed through our biological and cultural histories. Isn't it at least possible that there are aspects of reality that just won't be graspable in terms of human sized concepts?

“What is closed to the mind of a rat may be open to the mind of a monkey, and what is open to us may be closed to the monkey...armadillo minds cannot solve problems of elementary arithmetic but human minds can.” (McGinn 349-366)

What reason is there to think that our minds are capable of answering all possible questions? Might there not be problems beyond the reach of our minds as well. Up to that point the researchers of the paper are with McGinn all the way, there seems no reason to think that our minds will be able to comprehend absolutely everything. Could the link between consciousness and the physical be something we just aren't able to comprehend despite or because of the fact that consciousness seems so close. It is possible. What McGinn wants to argue however is that it is not just possible but actual, that an understanding of the link really is inaccessible to us. Here the argument goes something like this. What we are looking for is the length between subjective experience on the one side and physical reality on the other. But consider the two sides, our concepts regarding the subjective side come from introspection, from reflecting on our own phenomenal experience. Our concepts regarding the objective physical side come from our perception of an external world, independent of us. And perhaps inferences based on those concepts of perception. What kinds of concepts could we possibly use to understand the link between the two? It cannot be the introspective concepts because those are limited to the subjective side. It cannot be the objective physical concepts because those are limited to the objective side. Those are the only concepts we have and there is none of the concepts that can handle the link between the two sides. The link between consciousness and physical world will forever be beyond our conceptual reach. How good is that as an impossibility argument? It is not very good. The argument relies entirely on McGinn's dichotomous picture of our concepts, that they are all either internally introspective or externally perceptual and *never the twain shall meet*.

Our intellectual history is a history of astounding conceptual innovation. Think of the conceptual leaps of an Aristotle or a Galileo or a Leonardo or a Freud; consider the amazing range of concepts people work with. Concepts of spiritual enlightenment, prime numbers or

ethical obligations, of orders of infinity, of quarks and electrons, of physical universe of consciousness itself, many of our concepts do not fit neatly in McGinn's categories, as they are far from clear, incapable concepts that link the two. It certainly is possible that consciousness is beyond our conceptual grasp. But the researchers don't think McGinn has given a sufficient reason to believe that possible conceptual limitation is an actual limit. If we step back from Nagel and McGinn's arguments, then there is more to be said about anti-theories of this type.

An important effort of science is to try to figure things out. Sometimes it looks like we are getting closer, sometimes theories crash on the rocks of new data, but the effort to try to do things better and to get a better understanding, perhaps to get closer to reality continues. Despite what people often tend to think, science is rarely a matter of conclusive proof. It is more often a question of how the evidence leans at the moment or what theories seem most plausible in light of currently available evidence. It is also often a question of burden of proof as in a court of law. Is the theory innocent until proven guilty or guilty until proven innocent? Science often operates with guiding presumptions or working assumptions. Scientists operate as if theories are on the right track until they find reason to doubt them. To that quick sketch, the researchers want to add this consideration; sometimes theories do more than merely paint a picture of reality. Some theories have particularly strong implications for scientific practice itself. One such theory is a theory that says that the only possible answer to a particular question has to be this kind of answer; another kind of theory with particularly strong implications for scientific practices, one that says there is no possible answer to the question, give up.

Conclusion

The first kind of theory tells us that a research has to go to a certain direction; the second kind of theory tells you that research on a particular question is hopeless. The researchers think we should be particularly suspicious of those kinds of theories, whatever they are about. The basic reason to go about in such a way is because the implications for research are very strong. If we believe the first kind of the theory, the theory that says the only possible answer is this kind of answer, we will cut off any lines of investigations or try anything different. The result could be disastrous. If we are wrong we won't just have an incorrect theory, we will have stifled precisely the research that would have given us a better theory. The authors propose that we should come up with a presumption against that kind of theory just because the consequences for

research itself are so devastating if we happen to be wrong. The authors think that we should come with a similar presumption against the other kind of theory, the theory that says there is no possible answer to this question. It is because if we believed it, we would give up and if we give up, we will never find an answer even if we are wrong and there is really is one. That kind of scientific hypothesis is an inquiry stopper. The hypothesis itself is an obstacle for further research and that alone is a reason to come with a bias against it. In looking at anti-theories like Nagel's and McGinn's we might have asked two questions, one question is might the anti-theories be true? Might it be that we could never get an explanation of consciousness. A second question is, should we believe that the anti-theories are true? Should we believe that we will never get an explanation of consciousness? The researchers think the answer is different as it is quite possible that the anti-theories be true. It is also possible that we could never an explanation of consciousness. However, the second question wonders, should we believe that the anti-theories are true; not until our backs are absolutely forced to the wall. Should we believe that we will never get an explanation of consciousness: absolutely not. If we did believe that and conducted our research accordingly we would guarantee we will never get an explanation even if there really is one, just waiting for us but currently beyond our reach.

Works Cited

Bermudez, Jose Luis. *The Paradox of Self-Consciousness*. The MIT Press, 2000.

Cavanna, Andrea Eugenio, and Andrea Nani. *Consciousness: Theories in Neuroscience and Philosophy of Mind*. Springer, 2014.

Chalmers, David J. *Constructing the World*. Oxford University Press, 2012.

Costa, Claudio. *Lines of Thought: Rethinking Philosophical Assumptions*. Cambridge Scholars, 2014.

Hale, Bob. *Necessary Beings: An Essay on Ontology, Modality, & the Relations Between Them*. Oxford University Press, 2013.

Hartman, Geoffrey, and Daniel T. O'Hara. *The Geoffrey Hartman Reader*. Fordham University Press, 2004.

Lacewing, Michael. *Philosophy for A2: Key Themes in Philosophy*. Routledge, 2010.

Mandik, Pete. *Key Terms in Philosophy of Mind*. Continuum, 2010.

McGinn, Colin. "Can We Solve the Mind--Body Problem?" *Mind*, vol. 98, no. 391, 1989, pp. 349–366. JSTOR, www.jstor.org/stable/2254848.

Murtagh, Andrew, and Adam Lee. *Meta: On God, the Big Questions, and the Just City*. Cascade, 2017.

Nagel, Thomas. *The View From Nowhere*. Oxford University Press, 1986.

Penrose, Roger. *Shadows of the Mind: A Search for the Missing Science of Consciousness*. Oxford University Press, 1994.

Rosenthal, David M. *Consciousness and Mind*. Oxford University Press, 2005.