RATIONAL Realm

Exploring a rational approach to knowledge and life

Hoffman’s Conscious Realism: A Critical Review

Leslie Allan

Published online: 20 May 2022

Neuroscientist Donald Hoffman proposed a bold theory—that objects do not exist independently of us perceiving them and that all that really exists is conscious agents. In this critical review, Leslie Allan examines the three core components of Hoffman’s new idealism: Fitness Beats Truth (FBT) Theorem, Interface Theory of Perception (ITP) and Conscious Realism. Allan proposes solutions to linguistic absurdities suffered by Hoffman’s idealism before considering its most serious problems. These include oversimplifications and misunderstandings of evolutionary theory, self-refutation, heuristic sterility and dependence on the successes of scientific realism.

To cite this essay:

Allan, Leslie 2022. Hoffman’s Conscious Realism: A Critical Review, URL = <https://www.RationalRealm.com/philosophy/metaphysics/hoffman-conscious-realism.html>

To link to this essay:

www.RationalRealm.com/philosophy/metaphysics/hoffman-conscious-realism.html

1. Introduction

Professor Donald Hoffman

In a wealth of academic papers, videos and interviews spanning more than a decade, respected neuroscience researcher, Donald Hoffman, has proposed a novel solution to the mind-body problem. In the philosophy of mind, the mind-body problem specifically is the challenge of finding out what exists fundamentally; whether it is mind or physical bodies or some combination of both. Whatever we take to be ontologically fundamental, philosophers of mind are also tasked with explaining the nature of the relationship between mind and body. After more than two millennia of philosophers and scientists sweating over this problem, it appears no less intractable.

Hoffman's solution to the problem consists of the following three interconnected theories.

1. Fitness Beats Truth (FBT) Theorem
2. Interface Theory of Perception (ITP)
3. Conscious Realism

Working with other researchers, Hoffman's Fitness Beats Truth (FBT) Theorem posits that during the course of the evolution of species, organisms whose perceptual apparatus are tuned for fitness for reproduction always win out against organisms that are tuned to perceive reality accurately. The theorem purportedly results from mathematical modelling of the selective pressures operating during the evolutionary process. This leads Hoffman to propose a pictorial representation theory of sensory perception, named the Interface Theory of Perception (ITP). According to ITP, every organism sports a species-specific perceptual interface modelled on the metaphor of icons on a computer desktop. Just as icons on our computer desktop do not accurately mimic the underlying complex objects they represent, so do our perceptual representations of external physical objects hide their enormous complexity.

The combination of these two theories is consistent with a realist view of the external world; i.e., the view that physical objects and processes exist independently of minds that perceive them. It is with the third theory in Hoffman's tripartite synthesis that he recommends a radical departure from both common sense and the dominant scientific realist view of what actually exists. The first two posits provide the theoretical underpinning for Hoffman's Conscious Realism; the view that the real world consists solely of conscious agents. If FBT and ITP combine to show that we have good grounds for believing that our sense-perceptions of physical objects cannot possibly reveal their objective properties, then, for Hoffman, it's a short step to doing away with mind-independent physical objects altogether. For Hoffman, the key motivator for him in rejecting realism about physical objects is the scant progress made in solving the mind-body problem. His position is a form of philosophical idealism, in the tradition of Berkeley and Leibniz.

In this essay, I deal with each of Hoffman's three components in turn, highlighting what I think are some of the key problems. He uses his FBT and ITP theories to try and bolster his more radical metaphysical Conscious Realism thesis. So I'll start with examining his FBT and then move on to his ITP. Finally, I'll critically review his Conscious Realism thesis.

2. Fitness Beats Truth (FBT) Theorem

2.1 Modelling Natural Selection

Hoffman's form of metaphysical idealism, Conscious Realism, relies on an empirical-mathematical theory he and his collaborators term the Fitness Beats Truth (FBT) Theorem. Hoffman [2018: 1] summarizes this theorem like this:

> ... analysis of perceptual evolution using evolutionary game theory reveals that veridical perceptions are generically driven to extinction by equally complex non-veridical perceptions that are tuned to the relevant fitness functions. Veridical perceptions are not, in general, favored by natural selection.

In their paper illustrating their mathematical modelling, Hoffman and his collaborators reject the dominant view among evolutionary biologists by insisting that 'attempting to estimate the "true" state of the world corresponding to a given a sensory state, confers no evolutionary benefit whatsoever' [Prakasha *et al* 2017: 24].

I think that Hoffman and his team may have overstated their case here in a couple of ways. Their FBT modelling does not show all that they say it shows. It may show that in the long run fitness beats truth most of the time, but evolution is an ongoing process. So, in the short run, with constant genetic variation resulting from spontaneous mutations and the random pairing of alleles during sexual reproduction, some members of a species will sense truth to at least some extent.

Secondly, even in those cases in which fitness is maximized through selection, there remains some mapping to external reality, even if it is not homomorphic. Otherwise, to use Hoffman and his team's example [Prakasha *et al* 2017: 9] of organisms competing for a water resource, the organism using the 'Fitness-only' strategy won't get the water it needs to survive. Shermer makes exactly this same point when he writes, 'Finally, why present this problem as an either-or choice between fitness and truth?' [Shermer 2015]. Other writers have also picked up this same point, including Cohen [2015] and Vlerick [2014: 66f] summing up this more nuanced view of natural selection processes.

In fact, in an interview with Frohlich [2019] Hoffman seems to concede this point that our perceptions are truth-tracking to some extent. In replying to the question of why it is that we see the Milky Way when that perception has no fitness payoff, Hoffman says: 'So the idea will be that evolution has shaped us with a very simplified interface that's been shaped *mostly* [emphasis added] to report the stuff that's going to keep us alive'. Hoffman goes on to say that even so, we misjudge the relative distances to distant mountains, stars and the moon because these distant objects all have the same low 'fitness payoff' for us. So, according to Hoffman, our perceptual systems are 'representing' objective facts about fitness cost and 'report' information to us about objective reality to the extent needed to keep us alive.

More particularly, from the perspective of the science of evolutionary biology, another objection concerns how the FBT Theorem deals with the phenomenon in nature of mimicry. As Dickinson puts the objection:

> Hoffman would argue we see an icon that represents a snake, not a snake. But then why do nonpoisonous snakes evolve colorings to match poisonous ones? If there is no objective reality to mimic, why would mimicry prove a useful adaptation, and why would the interfaces of multiple species be fooled by such tricks?
>
> [Dickinson 2019]

Shermer [2015] makes the same objection. I'm inclined to think that Hoffman can answer this kind of objection within his framework without much effort. Could he not respond that what the nonpoisonous snake mimics is not the poisonous colourings of a real snake, but the attributes of the perception of the predator when that predator 'perceives' a poisonous snake? An analogy here is how a character in a virtual reality game can take on the appearance of another avatar in the game when such mimicry is to that character's advantage in the game.

Mausfeld [2015] advances a more general objection to Hoffman's approach. He finds serious problems with Hoffman's assumption that the notions of 'objective reality' and 'truth' have a role to play in biological theories of perception. Hoffman [2018: 11] responds to this objection by distinguishing between 'proximate' questions about internal mechanisms and 'ultimate' questions about the evolutionary development of those mechanisms. I find Hoffman's answer here a satisfactory response to Mausfeld's objection. Mausfeld also advances a number of conceptual objections to Hoffman's FBT Theorem and takes issue with Hoffman's use of selective constraints in the development of perceptual systems. I think these objections remain unaddressed.

Another objection is that in their mathematical simulations, Hoffman *et al*'s methodology rigs the game in favour of the strict interface strategy. Using Monte Carlo simulations, Hoffman *et al* [2015a: 1486] try to show that natural selection favours a strict interface strategy over naïve realist and critical realist strategies (shown in their *Fig. 1* and *Fig. 2* reproduced below). However, they are only able to demonstrate that conclusion using the contrived parameters they fed into their simulation. Hoffman and his team had set each player to perceive only four colours across the whole range of actual resource quantities spanning zero to 100. For the critical realist player, they perceive the colour blue when encountering actual resource quantities between 75 and 100. This methodological choice splits the resource quantities with the maximum payoffs across two perceived colours (yellow and green), increasing the realist players' cognitive load compared with their strict interface competitors.

Now note how if Hoffman and his team had set the perception of blue for both realist strategists to coincide with a resource quantity between, say, 40 and 60, then those players would be as finely tuned as the players using the strict interface strategy. Hoffman *et al* choosing to break the perceived colours for the realist strategy at precisely the mid-point of the Gaussian payoff curve was an entirely arbitrary choice that just happened to suit Hoffman *et al*'s FBT hypothesis. And that handicap for the realist strategies results from their methodologically arbitrary choice in carving up the resource map with an even number of colours (i.e., ‘where the perceptions of each player are limited to just four colors’ [2015a: 1486]). If Hoffman and his team had chosen an odd number of colours to input into their mathematical simulation, such as 3, 5, or 7, a single colour would have ranged over the resource quantities with the highest payoffs for both the realist strategies and the strict interface strategy. In Hoffman *et al*’s game, the strict interface strategy was guaranteed a win because they had tied one hand behind the back of the realist strategists.

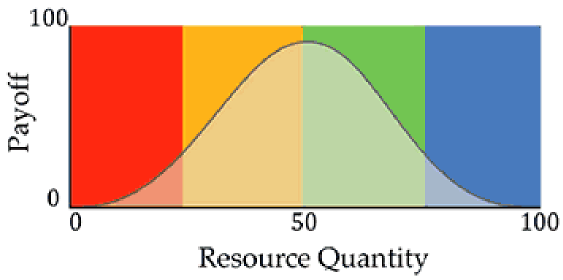

Fig. 2 A critical realist. The payoff function is approximately Gaussian. Any resource quantity between 75 and 100 maps to blue

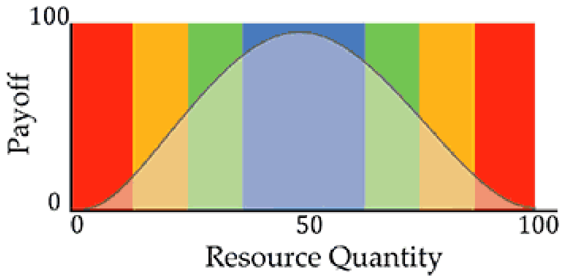

Fig. 3 An interface strategy. The resource quantities with the highest payoffs map to blue, and those with the lowest payoffs to red

Martínez [2019] mounts another serious objection showing how Hoffman *et al*’s mathematical modelling is overly simplistic. Here, he provides a number of empirical counterexamples to Hoffman’s thesis that fitness always beats truth. He shows that Hoffman’s simple model is accurate in only those cases in which an organism receives information from one source. Where an organism makes judgments based on two independent sources of information to ascertain utility, sending and receiving truthful signals promotes fitness. In the next section, I will explore how Hoffman *et al*’s oversimplification creates even more conceptual problems for their FBT theory.

2.2 Self Refuting

Let's assume that Hoffman *et al* [2015a: 1499] are right in thinking that 'all perceptions are fundamentally non-veridical'. Even with this concession, Hoffman *et al*'s Fitness Beats Truth (FBT) Theorem just seems self-refuting. Consider this. Hoffman is the result of human evolution and yet, purportedly, he discerns many facts grounded in observation.

Here is just one such fact about the world that Hoffman discovered from his observations of evolutionary processes and from working with his peers. He states that the classic argument that veridical perception offers a competitive advantage 'misunderstands the fundamental fact about evolution, which is that it's about fitness functions — mathematical functions that describe how well a given strategy achieves the goals of survival and reproduction' [Gefter 2016]. If we can't know anything about a world external to conscious minds undergoing selective environmental pressures, then how can Hoffman have found out that fact about just such a world?

Now, Hoffman could respond here that no observations were involved in these mathematical simulations. However, note that Hoffman did not make the required mathematical calculations in his mind, or even on pencil and paper. The mathematical models were programmed into physical computers and the calculations performed by same. He also communicated with his colleagues using physical telephones, exchanged ideas via written papers, and so on. All of these activities involved a myriad of everyday observations using his sense organs. If all of these observations are non-veridical, as he claims, then on what basis can he propose his supposed 'fact about evolution'?

Or take this other fact that Hoffman supposedly knows about the world:

> The environment in which our species evolved is a highly structured place, containing many regularities. Light tends to come from overhead, there is a prevalence of symmetric structures, objects tend to be compact and composed of parts that are largely convex, and so on. Over the course of evolution, such regularities have been internalized by the visual system (Feldman, 2013; Geisler, 2008; Shepard, 1994).

[Hoffman *et al* 2015a: 1490]

Again, this raises the question that if the operation of Hoffman's eyes and visual system are always non-veridical, as he claims, then how can he know so much about how they evolved and how they function? Hoffman seems to want his cake and eat it too.

The upshot here is that if Hoffman's perception of an external world populated by biological organisms undergoing selective pressures is all 'non-veridical', then so is any Hoffmanian theory built on it, no matter how elegant the mathematical models he produces. Here, Hoffman faces the horns of a dilemma of his own making. Either the observational data he relies on to produce his theory is veridical or it is not. If it is veridical, his FBT theory is false. In this case, his observational data is the case that breaks his FBT rule. On the other hand, if his observational data is non-veridical, his FBT theory is false as it is based on false premises.

Murphy [2020] makes exactly this point in his little piece when he writes:

> Hoffman does and does not believe that there's a reality as it is. He believes that there is a reality as it is when he discusses evolutionary theory/biology (i.e., when discussing our ancestors). And he doesn't believe there is a reality as it is when it comes to his philosophical position of conscious realism.

Hoffman and Prakash [2014: 17] make an early attempt to get around this objection that their conclusion that all of perception is non-veridical is self-defeating. They write in response:

> We claim that perception evolved by natural selection. Call this statement *E*. Now *E* is indeed informed by the results of experiments, and thus by our perceptions. We observe, from evolutionary game theory, that one mathematical prediction of *E* is that natural selection generically drives true perceptions to extinction when they compete with perceptions tuned to fitness.
>
> Suppose *E* is true. Then our perceptions evolved by natural selection. This logically entails that our perceptions are generically about fitness rather than truth. Is this a contradiction? Not at all. It is a scientific hypothesis that makes testable predictions.

The problem with Hoffman and Prakash's response here is that whether the conclusion, 'Natural selection generically drives true perceptions to extinction', makes predictions or not is wholly irrelevant to whether relying on perceptual evidence to support *E* contradicts the conclusion.

The following is a mutually inconsistent set of propositions.

D Perception supports *E*

E Perception evolved by natural selection

F Natural selection generically drives true perceptions to extinction

Hoffman and Prakash concede *D*. Now, even if we grant that we know *F* is true without using our perceptual capacities, we still need to remove at least *D* or *F* from the set to make it a consistent set.

Also, there is a certain irony in Hoffman and Prakash relying on what they say are non-veridical perceptions to test the truth of a scientific hypothesis.

In a later paper, Hoffman *et al* [2015a: 1500] try again to get around this objection. Here, they reply that their mathematical proof that natural selection does not favour veridical perceptions 'does not entail that all cognitive faculties are not reliable' and then proceed to defend the hypothesis that perhaps our ability to do maths and logic is accurate. However, our competence in maths and logic is beside the point. Hoffman *et al*'s mathematical simulations manipulate the conditions for the evolution of organisms in a natural physical environment. Their mathematical games have content. They have a subject matter. If the subject matter of those games (particular organisms with particular properties existing in particular environments) is not in actuality what they say it is, then no amount of mathematical manipulation will lead to anything other than erroneous results. If we can add flawlessly two and two to deduce four, that is of no help in adding two apples to two apples to arrive at four apples if the apples are not really apples, but oranges. As the saying goes, 'garbage in, garbage out'.

Later again, Hoffman [2018: 10] tries a different tack in answering a similar objection. As stated [Objection 4], the objection is that Hoffman's FBT Theorem assumes the truth of the theory of biological evolution with all its attendant biological entities, which are the self-same physical entities he is trying to prove that we know nothing about. Here again, Hoffman [Reply 4] defaults to pointing to the 'algorithmic core' of evolutionary theory, devoid of its ontological commitments. As he writes, 'Our theories are ladders to new levels of understanding, and sometimes a new level of understanding leads us to kick away the very ladder that led to it.' However, on that ladder is the very engine of the evolutionary process. For Hoffman [2018: 11], his FBT Theorem 'entails that DNA does not exist when it is not perceived'. In kicking away the very entities that form the ontological core of species' evolution (organisms, DNA, food resources, and so on), he leaves evolution a process without a subject. If Hoffman ends by denying the subject matter of evolution, what is there for the processes of reproduction and natural selection to work on?

2.3 Reductio Ad Absurdum

One way to view Hoffman's Fitness Beats Truth (FBT) Theorem critique of 'metaphysical realism', the view that there are mind-independent physical objects, is as a *reductio ad absurdum* form of argument. On this way of looking at Hoffman's approach, he is arguing that if the theory of evolution by natural selection is true, then nothing we perceive is veridical. A fundamental problem with this approach is that even if you assume it is true that all our perceptions are non-veridical, that fact would not entail that 'metaphysical realism' is false. That fact only entails that if such mind-independent objects exist, we can't faithfully sense their properties. As I have tried to show in this section, Hoffman's argument, in fact, presupposes the existence of such mind-independent objects for it to work; that is, just those objects that provide the selective pressures for fitness to evolve and to evolve in.

As a *reductio ad absurdum*, Hoffman needs the argument to stop at proving the absurdity of belief in the veridicality of perception. But even here, it does not do the work of disproving 'metaphysical realism'. Unfortunately for Hoffman, taking his *reductio ad absurdum* to its natural conclusion, it undercuts the very theory of evolution that he relies on to show that our perceptions are non-veridical.

As a further illustration of this point, consider Hoffman *et al*'s description of the requirements for setting up and testing their FBT Theorem using computer simulations:

> ... we must include strategies that see none of the true facts, some of the true facts, and all of the true facts. Even if we suppose that human perception is veridical today, we must consider all possible strategies, veridical or not, in order to explore the plausible hypothesis that we evolved from species whose perceptions were not veridical.

[Hoffman *et al* 2015a: 1482]

In this scenario, it is Hoffman and his collaborators who are evaluating whether the test subjects see the 'true facts' or not. Hoffman, though, is a member of the species *H. sapiens*, a species he says for whom today 'none of our perceptual experiences are literally true of the world' [Hoffman *et al* 2015a: 1489]. Again, his own thesis is reduced to a *reductio ad absurdum*.

I want to dig down a little deeper here with a couple of specific examples illustrating the absurdity of Hoffman's position. Consider again Hoffman's case study using an organism's perception of the quantity of a resource [Hoffman *et al* 2015a: 1486]. In their simulation, an organism employing the critical realist (veridical) strategy is pitted against an organism using the interface (non-veridical) strategy. Both types of organism perceive a variable quantity of a resource using just four colours. I've reproduced Hoffman *et al*'s illustration below (labelled *Fig. 2* and *Fig. 3* in Hoffman *et al*). One such resource is water. Now picture yourself sitting in front of a very large transparent tank that is being gravity-filled with water at a constant rate. We know that the tank is being filled at a constant rate because the force of gravity is constant and the water is being fed through a fixed diameter tube.

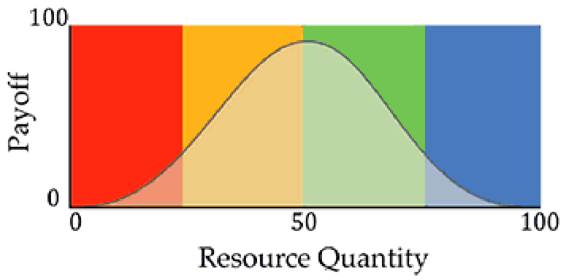

Fig. 2 A critical realist. The payoff function is approximately Gaussian. Any resource quantity between 75 and 100 maps to blue

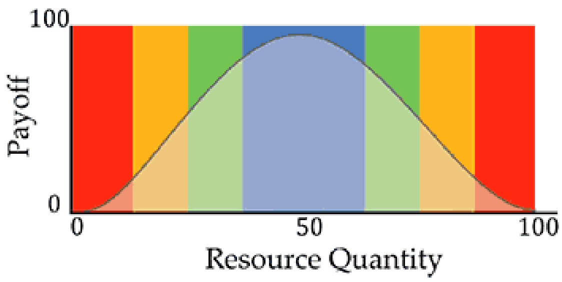

Fig. 3 An interface strategy. The resource quantities with the highest payoffs map to blue, and those with the lowest payoffs to red

Now Hoffman would have us believe that although the volume of water in the tank appears to be increasing, it may in fact not be so. For Hoffman, when the tank appears to us human beings using the non-veridical interface strategy as being two thirds full, it could in reality be one third full or it could be two thirds full (green bands in *Fig. 3*). You just don't know. The quantity you think you are perceiving is a 'you know not what'. Even more absurd, for Hoffman, when you perceive the tank as empty, it could in reality be full (red bands in *Fig. 3*).

Hoffman then applies this absurd conclusion to all of the other properties we sense. These include temperature, shape, velocity, hardness, and so on. If we cannot in reality determine the scalar quantities of these properties, then how do Hoffman and other scientists practice science? If a hydrologist measures the volume of water as, say, 70 cubic litres, and we believe Hoffman, the volume could be anything. It could be one cubic litre or it could be 1000 cubic litres. We just don't know. How then can we support a science of hydrology, or cosmology or biology at all?

Now, if Hoffman and other scientists can't say anything about the properties of biological organisms and the environments in which they live, then what are we to make of the entities that populate all versions of evolutionary theory? If cells, DNA, enzymes, fossils, mating pairs, predators, food, water, and so on, are all 'you know not what', if they are all non-veridical 'icons' representing 'you know not what' properties, then Hoffman needs to tell us how we can construct a theory of evolution in the first place. Hoffman appears to hoist himself on his own petard. Alternatively, if he insists on using the science of evolutionary biology to mount his case, he has already rejected the proposition that all perceptions are non-veridical. These are the horns of the dilemma that Hoffman is cast upon.

How did Hoffman get to this point of absurdity? One reason is that his research team's mathematical model is based on a highly simplified organism that experiences only four possible percepts representing a wide scalar range of resource quantities. Hoffman mentions one such resource that organisms need to survive: 'Not enough water and one dies of thirst; too much and one drowns; somewhere in between is just right' [Hoffman *et al* 2015a: 1486]. This is the kind of organism that would have first evolved some billions of years ago. Now compare that perceptual capacity to that of *Homo sapiens* today.

Hoffman assumes here without argument that the simple organism employing the useful strict interface strategy and with a capacity for only four percepts will remain unchanged over the subsequent billions of years of evolution. Contrast that simple organism with a current *Homo sapiens*' brain containing some 80 billion neurons with some 100 trillion connections between them. We are able to discern by several orders of magnitude more differences of quantities of water than just four. Hoffman's assumption that his team's highly simplistic mathematical simulation incorporates all of the strategy generation and selection forces guiding evolution throughout the last few billion years seems entirely fanciful.

In a later paper, Hoffman [2009: 20] recognizes the simplicity of a similar mathematical model and suggests how the models may be developed to more accurately reflect natural ecosystems. At one point, tellingly, he opines that these more complex models will allow scientists to determine at what level of complexity a truth strategy will prove more advantageous than a simple interface strategy. He bets, however, that there will be none of interest. For Hoffman, the jury is still out on this crucial question. In view of how this admission severely limits the confidence in his bold conclusions, Hoffman does not tell us his reasons for betting against what I would have thought the more reasonable outcome.

A second reason Hoffman got to this point of absurdity is that he and his team's mathematical model premises a direct relation between perception and action. This direct relationship was true for the earliest and simplest organisms in our evolutionary history. However, it is clear that during the long evolutionary process, this basic stimulus-response function was complemented with both an environment modelling function and an executive function. So, today, when we are hungry and sense an apple, we can decide to eat it now or later. Our decision depends both on the model of the external world we construct and on our desires and values. If our belief is that eating the apple will raise the chagrin of the person who owns it, we may decide not to eat it all. Now, none of these complexities about our biological capacities and their evolutionary origins figures in Hoffman's mathematical model. He simply assumes, against the evidence, that if organisms were simple one-directional stimulus-response mechanisms billions of years ago, they must essentially be the same now. In an informative paper, Vlerick [2014] has more to say on the complexities of our cognitive capacities and their evolution.

2.4 Time and Space

In the previous section, I highlighted a fundamental internal incoherence in Hoffman's advancement of his Fitness Beats Truth (FBT) theorem. In this section, I continue this theme by focusing, in particular, on problems with his notion that even our perceptions of space and time are non-veridical. Here, Hoffman tells us [Reply 4] that 'It is this algorithmic core that is used by the FBT Theorem to conclude that our perceptions of space-time and physical objects do not reflect objective reality' [Hoffman 2018: 10].

Let me start with Hoffman's conception of time. The first thing to note is that Hoffman's 'algorithmic core' is based on Dawkin's 'universal Darwinism' [2018: 10]. And this 'universal Darwinism' conceptually has as its object 'any group of entities that undergo transformations in terms of a change in probabilities between generations or iterations' [Campbell 2016].

Now, 'transformations', 'change', 'generations' and 'iterations' necessarily occur in time. In addition, central to the concept of 'evolution' is change over time. So, on both these counts, the very axioms of algorithmic Universal Darwinism used by Hoffman are grounded in the actuality of time and duration.

If the Universal Darwinism algorithms used by Hoffman, in combination with other assumptions, entail that one of the central axioms of Universal Darwinism is false, then either his acceptance of Universal Darwinism is misplaced or his conclusion that there is no time is false. Hoffman can't have it both ways. Undeterred by this logic, in the formal axiomatic presentation of his Conscious Realism theory, he readily and unblinkingly relies on the concept of time throughout his entire treatise. For example, he writes, 'The idea is that at each step of the dynamics each of the four kernels acts simultaneously and independently of the others to transition the state ... to the next state' [Hoffman and Prakash 2014: 10].

In writing on the evolution of the human eye, he similarly relies on the concept of time:

> A backward retina, for instance, with photoreceptors hidden behind neurons and blood vessels, is not the "best" solution *simpliciter* to the problem of transducing light but, at a specific time in the phylogenetic path of *H. sapiens*, it might have been the best solution given the biological structures then available.
>
> [Hoffman 2009: 2]

For Hoffman, the situation is no better with his rejection of the veridicality of our perception of space. Even if we grant that we know nothing of mind-independent objects, our category of space seems real. For example, within our visual field, that content of our visual experience we call our 'left hand' is to the left of our 'right hand'. What we call the 'Sun' in our visual field rises above what we call the 'horizon'. Even for Hoffman, when he draws upon his metaphor of physical objects as 'icons' on our computer 'desktop', he writes how 'a file icon is dragged to the trash' [Hoffman 2009: 13]. These are all spatial relationships integral to the way our perceptual experiences manifest to us.

3. Interface Theory of Perception (ITP)

3.1 What's Good about ITP

In the previous section, I examined the shortcomings of Hoffman *et al*'s attempt to show that the mathematical simulation of evolutionary processes proves that veridical perception is beaten to extinction in favour of a strict interface strategy. In that section, I dealt with some of the methodological and conceptual challenges facing how the Fitness Beats Truth (FBT) simulation is framed. Hoffman's Interface Theory of Perception (ITP) is a natural corollary of his FBT Theorem as it spells out explicitly the relation between what is perceived and the perceiver. His ITP extrapolates beyond the perceptual inabilities of the highly simplified players in his team's mathematical games to the limitations of highly developed creatures like you and me. To illustrate his ITP, Hoffman relies on a computer analogy. He and his team put it like this:

> Suppose that there is a blue rectangular icon in the upper right corner of the desktop for a text file that you are editing. Does this mean that the text file itself is blue, rectangular, or in the upper right corner of the laptop? Of course not. Anyone who thinks so misunderstands the purpose of the desktop interface. No features of the icon are identifiable with any features of the file in the computer. Moreover, one would be hard pressed to find a natural sense in which the icon is a veridical representation of the file. However, the icon is intended to guide useful behaviors. If, for instance, you drag the blue icon to the trash you can delete the text file; if you drag it to the icon for an external drive, you can copy the file.
>
> [Hoffman *et al* 2015a: 1484]

Even though Hoffman's ITP is underpinned by the game theory used in his FBT Theorem and thus stands or falls with it, his use of the computer analogy has come in for separate criticism. It is to those objections that I will turn to in this section.

Let me start by saying that I have a lot of sympathy with Hoffman's ITP, as do many, if not most, neuroscientists. What he writes about how our perception of the physical world is mediated is not remarkable given what neuroscientists have known for several decades about the neural cognitive processes behind illusions and the top-down processing of sensory information. Reflecting this dominant line of research, I think prominent neuroscientist Metzinger puts it well when he writes:

> ... a fruitful way of looking at the human brain, therefore, is as a system which, even in ordinary waking states, constantly hallucinates at the world, as a system that constantly lets its internal autonomous simulational dynamics collide with the ongoing flow of sensory input, vigorously dreaming at the world and thereby generating the content of phenomenal experience
>
> [Metzinger 2003: 52]

One point that I will be driving home in this section is that the fact that our perceptions of the external world are mediated by a range of factors does not preclude them from being ‘veridical’ in any ordinary or technical sense of the term. Realism about external mind-independent objects cannot be so easily dismissed.

3.2 Are Our Perceptual 'Icons' Veridical?

Let me start with Hoffman *et al*'s desktop icon analogy. He draws a distinction between realist strategies that mimic the world and 'strict interface' strategies in which perception does not preserve any structures of the world. As he and his team write, the latter winning perceptual strategy is analogous to using the icon on our desktop: 'No features of the icon are identifiable with any features of the file in the computer' [Hoffman *et al* 2015a: 1484].

However, in dismissing veridicality of the icon on our desktop, Hoffman and his collaborators may be pressing their analogy too far. Clearly, the icon on our desktop is representing a file in some respects. It's a mistake to think that to represent the file with some level of veridicality, the icon needs to be identical with the file. The icon is, after all, a representation. A map can faithfully represent the terrain it maps without being identical to the terrain. Cohen [2015: 1515f] drives home exactly this point in more detail.

Consider my perception of the tree in my back yard. When I see the tree, I experience a tree percept with brown and green features in my visual field. For my tree 'icon' to represent the actual tree in my yard, it need not reveal all (or even most) of the features of the actual tree in my yard. We don't require my tree percept to reveal the cells, molecules and atoms that make up the actual tree for it to be veridical. We don't require my percept to represent the complex process of photosynthesis that is going on in the actual tree.

For my percept to be veridical, we require it to stand in the appropriate causal relation to the tree planted in my back yard. So, we require the actual tree to be an essential part of the causal chain of happenings that start from light from the sun reflecting off its surface particular wavelengths in the brown and green parts of the electromagnetic spectrum. The causal chain continues with those waves hitting the photoreceptors in my eyes, then with electrical impulses being transmitted along my optic nerves and then being processed in the visual cortex part of my brain. This final part of the causal chain leads me to have the private phenomenological experience of the tree. When these causal relations between the actual tree and my tree percept (the tree 'icon') are realized, we say that my perception of the tree is veridical.

In the same way, we say that the file icon on my desktop is veridical when it stands in the appropriate causal relation to the actual file in the computer. And by design, I create that causal connection when I configure the properties of a desktop icon to point to an actual file within the folder structure of the computer's storage. Now, when I drag the icon to the trash or move it to a different folder, the actual file is deleted or is moved. It is precisely when I drag the icon to the trash and it is not deleted, or when I move the icon to a different folder and the wrong file is moved, that we say that the icon does not represent that file; that the icon is not veridical.

A second consideration about Hoffman *et al*'s 'icon' analogy is that it reveals more than intended about the nature of realism. Just as we can peer behind the desktop icon to reveal the inner workings of the computer interface, likewise, we can dig behind our tree percept to uncover the inner workings of nature. With our desktop icon, we can investigate how moving the icon into another folder icon with our mouse changes the electrical patterns on the hard disc. Likewise, scientists investigate how the various colours of a perceived tree represent certain wavelengths of reflected light and how they interact with photoreceptors in our eyes and transmit electrical signals to our visual cortex. Our tree percept does not place us in an epistemological prison; no more so the icon on our desktop.

3.3 Perceiving Distance

Now, granted, there are many times when the items in our sensory field—our ‘icons’—don’t reflect reality as it is. But does this fact entail that our percepts are always non-veridical? Hoffman *et al* [2015a: 1497] point to ‘obvious cases where our perceptions radically disagree with our careful measurements’. They point to how the ‘sun, moon and stars, for instance, all look far away, but they all look about equally far away’. Citing the research of other neuroscientists, they add that ‘[e]ven at close distances our perceptions differ from our careful measurements’.

I’ll return to the case of perceiving distant objects shortly. But is it really the case that our perception of distance up close is always non-veridical? Hoffman *et al* [2015a: 1497] point to research studies that purportedly show this. In fact, they seem to show the opposite; that in many perceptual circumstances, our perception of distance is veridical. For example, Kappers [1999: 1001] studying haptic perception in subjects makes only the weaker conclusion that ‘haptic perception of, for example, distances or parallelity does not always conform to physical reality’. And where there is deviation from reality, Kappers found an algebraic relation between the two. As she concludes: ‘The results clearly show that what subjects feel as parallel deviates systematically from what is actually physically parallel [Kappers 1999: 1009]’. In Hoffman *et al*’s [2015a: 1484] mathematical simulations of the evolutionary process, such algebraic relations (either isomorphic or homomorphic) count as veridical within a realist strategy.

Koenderink *et al*’s [2010: 1163] study of the visual illusion resulting in perspective distortion cited by Hoffman *et al* [2015a: 1497] also reveals an algebraic relation (viz.: ‘large systematic deviations’) between perception and reality. And again, Pont *et al*’s [2011: 6] study of depth perception finds an algebraic relation in the subjects ‘adjusted foreshortenings as a function of distance and size’.

Now, these researchers sought to demonstrate how the particular perceptions that are the subject of their study are not ‘veridical’. However, note how their use of the term ‘veridical’ is used in the sense of there being an identity relation between the subject’s perception and reality. In contrast, Hoffman *et al* [2015a: 1483] define all of the versions of realism that they seek to model, from that of the ‘omniscient realist’ to that of the ‘critical realist’, as acting on either an isomorphic or the less stringent homomorphic mapping between perceptions and reality. It is a surprise, then, that Hoffman *et al* [2015a: 1497] end their citation of these studies with a wholehearted assent to Koenderink’s conclusion that the ‘very notion of veridicality itself . . . is void’ [Koenderink 2014: 5]. Perhaps it is not so surprising given that Koenderink is a Hoffman acolyte.

A further point worthy of note here is that the researchers cited examined perceptual distortions under unusual conditions (e.g., one eye blindfolded, objects hidden by a table). However, there are many circumstances in which such distortions are wholly absent. For example, when I lay out marbles evenly spaced at particular points along a ruler laid out on the floor, what reason is there for thinking that my perception of the space between each marble is non-veridical? Under these kinds of normal perceptual conditions, neuroscientists, in fact, have found in the mammalian brain neuronal grid cells whose firings map out just such evenly spaced locations in the environment. (See, for example, Balkenius and Gärdenfors 2016, McDermott 2020 and Moser *et al* 2008.) This linear mapping between perception and the environment is what we would expect if perception of local distance is veridical.

Let me now return to the case of perception of objects very far away, such as the moon and the stars. For these objects, Hoffman offers a standard evolutionary account for why we misperceive their relative distance. As he responds during an interview with Frohlich [2019]: ‘So the idea will be that evolution has shaped us with a very simplified interface that’s been shaped mostly to report the stuff that’s going to keep us alive.’ He explains further by contrasting the evolutionary advantage gained from perceiving close distances accurately:

> Space and time are just a data structure. They’re there to represent fitness payoffs. The distance from me to an apple, say, here like two meters away versus another apple, you know, 20 meters away—that distance is coding the percentage of my caloric resources that I currently have that would be required to be expended to get the resources in the apple at two meters versus 20 meters. In other words, distance is a calorie expenditure representing fitness cost. And so, it’s no surprise that the stars look about as far away as the mountain, because they’re both at infinity given my caloric resources.

[Frohlich 2019]

Here, Hoffman has let the cat out of the bag. To explain why perception of very large distances is non-veridical, Hoffman illustrates his point by contrasting this case with a situation in which veridical perception of distance is necessary for survival. In the case of getting an apple, evolution selects for accurate perception of distance, otherwise we would be expending uneconomical amounts of calories.

Hoffman could claw back here and simply insist that all perception of distance is non-veridical. This would not be a good move as it would render his and his team’s own explanation of their thesis incomprehensible. For example, I refer you to Hoffman *et al*’s [2015a] explanation of their FBT Theorem. To help us understand their thesis, they construct a number of explanatory diagrams. By the time we get to the diagrams, Hoffman and his team have already told us that using the strict interface strategy that we all do, ‘none of our perceptions reflect the structure of the world’ [2015a: 1484]. Our perceptual mappings to the world are neither isomorphic nor more liberally homomorphic, as they illustrate in their *Fig. 3* (reproduced below) [2015a: 1486]. So, they instruct us, for an object of perception that we perceive as green, the actual scalar quantity of that object could be around 30 or 70. We just don’t know and there is no way of knowing.

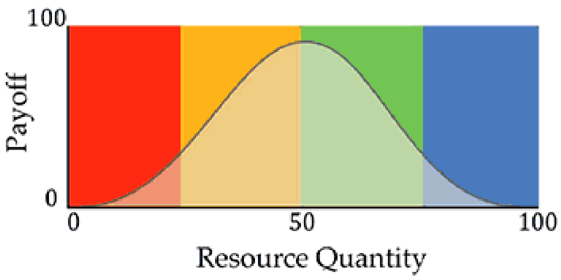

Fig. 2 A critical realist. The payoff function is approximately Gaussian. Any resource quantity between 75 and 100 maps to blue

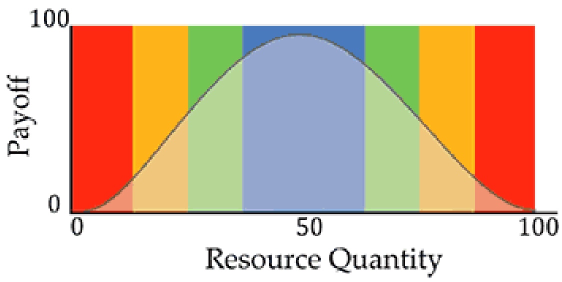

Fig. 3 An interface strategy. The resource quantities with the highest payoffs map to blue, and those with the lowest payoffs to red

Now, apply that revelation to how we should interpret *Fig. 2* on the same page (reproduced above). Look at the 'Resource Quantity' axis in Hoffman *et al*'s diagram. If our perception of distance preserves *no* structure of that to which we are looking, then when we look at the '50' resource quantity point that appears half way along the horizontal axis between the '0' and '100' anchors, that point could in actuality be mapped to either around 30 or 70. According to Hoffman *et al*'s ITP, we just don't know. If, *ex hypothesi*, colours do not map isomorphically or homomorphically to actual resource quantities and our perception of distance is just as non-veridical as perception of resource quantity, then our perception of points along a line likewise leaves us clueless about actual distance along a line. This is not just a matter of poor precision in sensing distance along Hoffman *et al*'s drawn line. If we are to believe their radical thesis, on their own terms, for each point on the line, we have no idea whether that point indicates a greater or lesser distance compared with the points to its left and to its right. This creates a problem not only for making sense of Hoffman *et al*'s diagrams, but also for our understanding of any chart that uses axis.

3.4 Parasitic on Realism

In this section, I want to advance the objection that Hoffman's Interface Theory of Perception (ITP) borrows from his opponents' metaphysical framework without paying its dues; that is, that his scheme is parasitic on realism. Consider the answer that Hoffman offered in response to an objection put to him. The objection he deals with is this:

> If natural selection did not design our senses and brain to construct a relatively accurate model of reality, then how is it we can land a spacecraft on Mars with pinpoint accuracy, or put satellites in space so finely tuned with relativity theory that they can detect a device on earth within less than a meter of accuracy?

[Hoffman 2018: 8]

Hoffman's reply to this objection makes it seem as if successfully landing a spacecraft on Mars is simply a matter of getting our hand-to-eye coordination right. As he puts it: 'one can have perceptions of 3D space and 1-D time together with perfectly coordinated actions in that perceived space and time that are entirely predictable to arbitrary accuracy' [2018: 8f]. Even though, for Hoffman, this ability 'entails absolutely nothing about the nature of objective reality', we manage to get around using that interface, much like the subject who manages to get around quite well after a while when fitted with upside down goggles that reverse the visual image. (See, for example, Abrahams [2012]).

Here, Hoffman invokes his and his team's Invention of Symmetry Theorem (IOS) to demonstrate mathematically how we can map, using symmetry, planarity and compactness, perceived shapes in a 3D space to another non-actualized dimension. Their system of perception-behaviour is highly operationalized: 'the regularities of our perceptions are an evolutionarily designed interface that guides adaptive behavior and hides the true regularities of the objective world' [Hoffman 2018: 9].

That mathematical mapping scheme may be adequate for explaining the successful manipulation of 'icons' within a spatio-temporal field. When it comes to landing a spacecraft on Mars, what Hoffman *et al* ignore is the essential role that theory construction plays in scientific achievements.

The feat of landing a spacecraft on a planet millions of kilometres from Earth was not achieved by simply using a joystick to manipulate the 'icon' of a spacecraft within a virtual reality game. That achievement is not like playing a space video game in which our only task is to co-ordinate our eyes and hands within a set of pre-established game rules. Landing a spacecraft on Mars is the end result of centuries of hypothesizing and testing models of reality in the fields of physics, astronomy and cosmology against predicted future perceptions. To put a spacecraft on Mars, scientists first needed to understand what lies under the bonnet of our everyday perceptual world, so to speak. Space scientists only got to know how to build the spacecraft, send it into space and control its landing by first theorizing and understanding the underlying physics that govern planets, rockets and spacecraft.

It is only after the theoretical and experimental hard work had been done that Hoffman *et al* are able to reverse engineer the mapping of manipulations of physical objects by scientists in space-time to achieve the Mars landing. What scientists did over many years was uncover the inner workings of nature so that they could predict how the Mars spacecraft would perform in hitherto unknown environments. What they did is much like a video gamer working out how the microprocessors, power supply, display, memory devices, and so on, interact with software code to display the visual image she sees in her headset. No amount of looking more closely at and manipulating the icons on her screen will allow her to predict with much success what will happen if she performs a motion that is not within her known game rules. For that, she must understand how the virtual reality apparatus really works and the programming rules that govern its behaviour.

Hoffman is well familiar with how scientists create and validate predictive models. In his interview with Tsakiris [2020], he speaks about one of those crucial understandings space scientists needed to have under their belt if they were to be successful in landing a spacecraft on Mars. Speaking of Einstein, he said:

> ... in 1905 and then in his general theory of relativity, he said, 'Here's my equation, I'm going to predict exactly how much light will bend and precisely where Mercury will be and we'll test that against Newton.' And it was when the experiments confirmed the precise mathematical predictions of his equation of general relativity, that's when he burst on the scene

Here, Einstein was not postulating about how some unknown substance behaves in some unknown substrate. He was theorizing in particular about how the mind-independent body we know as Mercury deflects light. Einstein was theorizing about how mass and energy behave in space and time independently of how we perceive them. That's why predictive success signals a theory's veracity and why Hoffman wants to emulate such success. Under Bayesian inference rules, the more novel the prediction (i.e., the more unlikely the predicted event will occur if the theory from which it is derived is false), the more epistemic weight it lends to the theory while reducing the probability of the rival theory. In this case, the stunning confirmation of the predicted value of the perihelion of Mercury provided independent evidence that mind-independent bodies behave as General Relativity describes.

Hoffman recognizes this objective feature of novel confirmations when he says in the same interview: 'once you write down a mathematical theory, the theory becomes smarter than the person who wrote it down'. In Einstein's case, as Hoffman relates: 'he had no idea that it was going to predict black holes. He didn't know that. But the equations do predict black holes, and he didn't like it'. Again, the prediction of hitherto undreamed of black holes pointed to new real objects out there. With Einstein's successful prediction of black holes, Hoffman tantalizingly seems to accept a realist interpretation of physical theories when he says: 'the theory had a deeper, in some sense a deeper insight into reality than the person who wrote it down'.

It is this deep insight into reality that our most successful physical theories afford us and that enable scientists to make incredible accomplishments. And it is this window into reality that, historically, realism has given us that Hoffman *et al*'s ITP fails to acknowledge in its reverse engineering of these successes. Their operationalization of the activities of scientists and *post hoc* cross-mapping of hand-eye co-ordination is reminiscent of the logical positivists and phenomenalists of the mid-twentieth century with their reconstructions of 'physical object' talk into 'permanent possibilities of sensation'. As with their predecessors' failed logical positivist and phenomenalist programme, their operationalized translations of scientists' successes is parasitic on the realist theories that enabled those achievements. (For more on parasitic phenomenalism, see my Allan [2016a].)

The ability of scientists to land a spacecraft on Mars raises another problem for Hoffman. He argues that evolution shaped our perception not to mirror truth, but to enhance our reproductive fitness. As he puts it: 'Perception is not about seeing truth, it's about having kids' [Hoffman 2018: 5]. If our perceptions are shaped by fitness payoffs—with opportunities to bear more children—then that raises a question: How does our ability to manipulate 'icons' with such precision to enable scientists to land a spacecraft on a body millions of kilometres from anyone aid reproductive fitness? Through advances in technology, scientists also now have the capacity to see molecules and atoms and to see distant galaxies. How does seeing these 'icons' help scientists have more children? *Prima facie*, such abilities appear to enable the opposite; a time and resource consuming distraction away from finding a mate and reproducing plentifully. In fact, surveys reveal that increasing education levels are a reliable predictor of reduced fertility [Roser 2017] and that female scientists abandon their career to have children [Elaine and Lincoln 2011; Else 2019]. This counter-intuition to evolutionary theory is explained easily on the standard view that *Homo sapiens*' ability to model reality co-evolved with the ability to find a mate and reproduce. On Hoffman's scheme, it is another mystery waiting to be solved.

4. Conscious Realism

4.1 Object Permanence

In the previous two sections, I reviewed Hoffman's Fitness Beats Truth (FBT) Theorem and his Interface Theory of Perception (ITP). Both these theories worked together in Hoffman's endeavour to undermine the realist view that our perceptions are at least sometimes veridical; that they tell us something about the world of mind-independent objects. With Hoffman's Conscious Realism thesis, we see the philosophical culmination of his FBT and ITP. This is the most radical of his ideas and the most difficult to swallow for almost everyone who comes across it.

After casting doubt on what we can know about a mind-independent reality, Hoffman's next step is to do away with it altogether. All there is are conscious agents. Hoffman summarizes his thesis thus:

> So, instead of proposing that particles in spacetime are fundamental, and somehow create consciousness when they form neurons and brains, I propose the reverse: consciousness is fundamental, and it creates spacetime and objects. ... reality is a vast social network of interacting "conscious agents," in which each agent has a range of possible experiences, and each agent can act to influence the experiences of other agents.

[Hoffman 2019b]

Hoffman explains how he was driven to this radical conclusion because of the lack of progress on the mind-body problem. As he puts it:

> To this day, science has not dispelled the mystery. Does neural activity cause conscious experiences? Some think so but have no idea how. No neural cause has been proposed for even one conscious experience. Precisely what neural activity causes, say, the taste of vanilla, and precisely how and why does it do so? No one knows.
>
> Are conscious experiences identical to, rather than caused by, neural activity? Some think so but again cannot give even one example. Precisely what neural activity is identical to the taste of vanilla? No one knows.
>
> Why has the hard problem of consciousness remained intractable for centuries despite determined efforts by brilliant scientists? I think the culprit is our assumption that our perceptions reveal a reality that exists even if unperceived.

[Hoffman 2019b]

Once we do away with that pestering assumption about mind-independent objects, Hoffman thinks, we get a clear view of how to proceed in our research. Hoffman's view here is characterized as falling under that type of ontology philosophers call 'idealism'. Berkeley, Spinoza and Whitehead are notable proponents of this kind of view. What is not evident is why Hoffman labelled his view 'Conscious Realism', adopting his opponents' 'realist' moniker. However he names his idealist theory, it is faced with some key problems. I'll begin this section by discussing some semantic issues with Hoffman's thesis and some suggestions for improving it. I'll then go on to more substantive issues, including his theory's heuristic sterility and reliance on realist assumptions.

The first challenge with Hoffman's Conscious Realism is that it reduces our ordinary way of speaking about everyday objects to absurdity. For Hoffman, physical objects literally pop into and out of existence as we change our point of view. Hoffman puts it this way: 'we create an apple when we look, and destroy it when we look away. Something exists when we don't look, but it isn't an apple, and is probably nothing like an apple' [Dickinson 2019].

But if the apple ceases to exist when I'm not looking at it, then how can something that ceases to exist nourish me once I've swallowed it and can no longer see or feel it? For that matter, what stops me from falling from the sky when I'm not looking at the aeroplane wings on my flight to London? Hoffman could say that the wings still exist because other passengers are looking out the window. Then what happens when all of the passengers look to the front of the plane, or fall asleep? What keeps the train moving on my journey to Sydney when no one is looking at the wheels? These are the same kinds of absurdities that Berkeley [1710 (1974)] faced with his version of ontological idealism. Berkeley's solution was to pass the job of keeping unperceived objects in mind onto God in order to guarantee that apples, aeroplane wings and train wheels didn't disappear when no one was observing them. However, this option is closed off for Hoffman as there is no place for God in his explanatory account.

One move that Hoffman can make here is this. Drawing on his virtual reality game analogy [Hoffman 2019b], when we are playing *Grand Theft Auto* and we look away from a competitor about to crash into our car, we don't say that the other car disappears in that moment. When we are *in* the simulation, it makes no sense to think that the objects we are interacting with disobey the laws of physics whenever we turn away. Even when we are out of the simulation—when we are speaking at a meta-level about the game—it makes no sense to speak of the competitor's car disappearing from the game whenever we looked away. Competitor cars and everything else in the game are designed to obey the laws of physics, which includes *not* disappearing when unobserved.

The same holds for apples and aeroplane wings in Hoffman's simulation. If objects in Hoffman's simulated world obey the simulated laws of physics, then they can't disappear when we are not observing them. If they did, then they are not the same 'apples' and 'aeroplane wings' that we are talking about in ordinary language when we refer to these things. So, Hoffman can avoid the absurdity of apples disappearing when unobserved by using a different language when he is referring to the simulation at the meta-level. When we speak outside of the game of *Grand Theft Auto*, we say something like this: the visual image of the competitor's car ceased to be generated when the player's attention was diverted. Likewise of the apple, to avoid absurdity, I suggest when Hoffman speaks at the meta-level that he say something along the lines that the appearance of the apple disappears when unobserved.

4.2 Object Individuation

A related problem for Hoffman is how to individuate objects, such as tennis balls and cars, on his schema. A realist view about external physical objects explains naturally the singular identity of things that are perceived by more than one person. Hoffman's theory, on the other hand, generates puzzles that strain language conventions to breaking point unnecessarily. For example, Hoffman writes:

> According to [multimode user interfaces] MUI theory, the objects of everyday experience – tables, chairs, mountains, moon – are not public. If, for instance, I hand you a glass of water, it is natural, but false, to assume that the glass I once held is the same glass you now hold. Instead, according to MUI theory, the glass I held was, when I observed it, an icon of my MUI, and the glass you now hold is, when you observe it, an icon of your MUI, and they are numerically distinct. There are two glasses of water, not one. And if a third person watches the transaction, there are three glasses.
>
> [Hoffman 2008: 97]

Hoffman wants to respond to the obvious objection by Searle that successful communication between us requires a public language in which, when we point to a common object, the word we use for that object must mean the same for both of us; that we are referring to the same object. Otherwise, we will be speaking past each other.

Hoffman answers by using the example of a virtual tennis game played by Bob and Tom. As Hoffman puts it:

> Bob and Tom, playing virtual tennis, can talk meaningfully about "the tennis ball" they hit; they can agree that Tom hit "the tennis ball" out of court, thus losing a point. There is, patently, no public tennis ball. Instead, a supercomputer in the back room feeds signals to the helmet displays of Bob and Tom and each, in consequence, constructs his own tennis-ball experience. But Bob's tennis-ball experience is numerically distinct from Tom's. And there is no other tennis ball around to serve the role of public tennis ball. Thus public physical objects are not required for meaningful communication.
>
> [Hoffman 2008: 97]

Hoffman is right in saying that Bob and Tom agree that they are playing with the same tennis ball. They do that in accordance with the linguistic rules we all commit to when playing networked virtual games. However, they are prevented from doing just that under Hoffman's proposed linguistic account. If we accept Hoffman's explanation of what is going on, Bob and Tom are playing with numerically different tennis balls (and in numerically different tennis courts). They can't then be playing against each other in the same game of tennis. On Hoffman's account, they must be playing two separate games of tennis. *Contra* Hoffman's claim, this numerical separation of tennis balls will actually mess up meaningful communication between people trying to co-ordinate their actions. Under Hoffman's scheme, when Bob hits the ball into Tom's part of the court and Tom misses, Bob can't justifiably claim a point. Tom will complain that the ball that he missed is not the same ball that Bob had hit. They are different balls. Likewise, if I steal your car, I haven't really stolen your car, have I? The car sitting in my driveway that looks identical to your car is not really yours. So, don't call the cops.

The linguistic entanglements only get worse. In the same paper, Hoffman writes:

> According to MUI theory, everyday objects such as tables, chairs and the moon exist only as experiences of conscious observers. The chair I experience only exists when I look, and the chair you experience only exists when you look.
>
> [Hoffman 2008: 98]

The implication here is that according to Hoffman, when I sit on a chair, I'm sitting on a conscious experience. And when I look at the moon through my telescope, I'm really looking at a conscious experience. As Hoffman explains: 'We only see the chair icons we each construct each time we look' [2008: 98].

The absurdities continue to mount. If a 'chair' is really a conscious experience, then, for Hoffman, when we see a chair, we are really seeing a conscious experience. There are really then two experiences going on: the 'chair' experience and the experience of seeing the 'chair' experience.

I think there is a way for Hoffman to avoid these kinds of linguistic absurdities. Getting back to Hoffman's virtual reality game analogy, I suggest he avoid saying that the tennis ball is a conscious experience. I suggest instead that he say that the percept of the tennis ball is a conscious experience. Think of the language rules we use when we are playing a virtual reality game. When we play virtual tennis, we don't say that we are hitting a conscious experience around the court and we don't say that we are seeing a conscious experience when we see the virtual tennis ball. Adopting my suggestion here allows Hoffman to say in real life, as in the virtual reality game, that we 'see the tennis ball' and that seeing the tennis ball consists in the having of the experience of a tennis ball.

Hoffman could avoid the other absurdity that we play with physics-defying magic disappearing tennis balls in real life and in virtual reality tennis games by appealing to the notion that tennis balls are theoretical fictional entities by which we take an imaginative leap beyond our immediate experience. This approach leverages off the realist conceptual framework positing mind-independent objects in order to preserve the fictional character of his virtual reality analogy. Borrowing from the realist conceptual scheme is an obvious disbenefit for Hoffman. However, I suggest this move is less debilitating for him compared with advocating absurd linguistic expressions.

With his heavy reliance on the virtual games analogy, it's difficult to see how Hoffman can escape presupposing a realist conceptual scheme. To develop this point further, consider this. In defending the notion that Bob and Tom are playing with numerically distinct tennis balls while still speaking of 'the tennis ball', Hoffman opines: 'And there is no other tennis ball around to serve the role of public tennis ball. Thus public physical objects are not required for meaningful communication' [Hoffman 2008: 97].

What we need to keep in mind here is that in the case of a virtual game, Bob and Tom both know that there are mind-independent physical structures and processes that generate the simulation of the 'public tennis ball'. Hoffman himself refers to how the 'supercomputer in the back room feeds signals to the helmet displays of Bob and Tom and each, in consequence, constructs his own tennis-ball experience' [2008: 97]. Hoffman could object here that Bob and Tom might not 'know' that there is such a supercomputer working in the background. Perhaps there is some other set of structures and processes in place that generate the tennis ball experience. However, it remains the case that it is Bob and Tom's *belief* that there is such a supercomputer that warrants their confidence that they are playing the same tennis game and not locked into their own solipsist world. Even when we acknowledge the point that Bob and Tom could be wrong in their belief, the crucial point here is that Bob and Tom must have *some* belief in a set of structures and processes that are independent of them, yet common to both of their worlds, that generates the image of the tennis ball. It is this shared belief in a common source of their tennis ball images that warrants their language convention; a shared belief that is precisely denied by Hoffman.

Now, Bob and Tom may be completely ignorant of supercomputers and how they process information and generate images in their headsets. Nonetheless, Bob and Tom accept that whatever the system is that generates the simulated tennis game, it is simulating the physical process that generates visual information and transmits it into their eyes. For Bob and Tom to accept that they are playing the same virtual game with the same tennis ball, they are assuming this second-order physical processing and generation of visual information that mimics the usual first-order perception of objects via mind-independent physical objects reflecting light-waves of particular wavelengths and intensities into their eyes.

In the absence of a worked out theory of a consciousness-only network that generates all of our perceptual experiences that rivals the explanatory and predictive power of our current realist theories of perception, Bob and Tom remain wedded to their realist assumptions underpinning their belief that they are playing the same tennis game with the same tennis ball. The upshot here is that Hoffman's use of the virtual reality tennis game analogy plunges him into linguistic muddles and absurdities that only serve to underscore how Bob and Tom can only play the virtual game by making realist assumptions about the underlying metaphysics.

4.3 Collapse to Solipsism

I now move on to more substantive objections to Hoffman's Conscious Realism. A particularly incisive observation that Dickinson [2019] makes is that if the icons on our desktop reveal nothing about a mind-independent reality, then perhaps our consciousness is just another icon. Dickinson puts the objection quite succinctly:

> If our perceptions of reality are merely species-specific interfaces overlaid upon reality, how do we know consciousness is not simply another such icon? Maybe the "I" of everyday experience is a useful fantasy adapted to benefit the survival and reproduction of the gene and not part of the operating system of reality.
>
> [Dickinson 2019]

This is a real thorn in the side for Hoffman's thesis as at its ontological core is the postulation of 'conscious agents' as distinct entities. Perhaps Dickinson is even being too generous to Hoffman here because it seems the conscious agent, our 'I', is not even an 'icon' on our desktop. As the Buddha and Hume have pointed out, we don't perceive the thing doing the thinking. The conscious agent, our 'I', is just as much a theoretical construct as the mind-independent physical objects we perceive.

But what of other conscious agents? The existence of other minds seems to be on even shakier foundations. Whatever can be said of perceiving our own mind, we don't perceive other minds at all. They are not even represented as 'icons' in our perceptual field. We only perceive the physical bodies that they purportedly inhabit. If, on Hoffman's view, we should ditch our belief in physical bodies in order solve the intractable mind-body problem, then how more so should we abandon belief in other conscious agents to solve the wicked problem of other minds? It's hard to see how Hoffman's metaphysical clean up does not lead us inexorably down the path to accepting a thorough-going solipsism. This is another of those problems faced by Berkeley [1710 (1974)] some three centuries ago resurfacing to bother modern idealist Hoffman.

Paradoxically, at first, Hoffman doubles down, claiming that our 'icons' of other conscious agents 'give deeper insight into the objective world' compared with that given by our 'icons' of inanimate physical objects [Hoffman 2008: 103]. A little further on, Hoffman does put in a half-hearted attempt to 'get outside of our epistemic jail, the super-user interface' to know about other minds [Hoffman 2008: 110]. For this, Hoffman suggests we look in the mirror.

> All you see is skin, hair, eyes, lips. But as you stand there, looking at yourself, you know first hand that the face you see in the mirror shows little of who you really are. It does not show your hopes, fears, beliefs, or desires. It does not show your consciousness.
>
> [Hoffman 2008: 110]

Next, Hoffman makes the crucial leap from recognizing first-hand one's own consciousness behind the appearance of one's face to recognizing the consciousness of others.

> All you see, and all that the user interfaces of others can see, is literally skin deep. Other people see a face, not the conscious agent that is your deeper reality. They can, of course, infer properties of you as a conscious agent from your facial expressions and your words; a smile and a laugh suggest certain conscious states, a frown and a cry others.
>
> [Hoffman 2008: 110]

The problem here is that we already know how we make the inference to other conscious agents through reasoning by analogy. The challenge for Hoffman is showing how this inference is reasonable given that these purported other minds are just as much hidden behind non-veridical icons as are physical objects. Hoffman jumps immediately from the first-person account to supposing what '[o]ther people' infer without giving us a reason for thinking that such other people exist. This is precisely the question at issue that Hoffman avoids in his argument. With due credit to Hoffman, he does concede that the inference to other minds is 'unavoidably fallible' [2008: 110]. None the less, if our 'user interface' hides reality, as Hoffman claims, then he hasn't really even begun to answer the question of his interlocutor: How do we get outside of our epistemic jail to know other minds exist? Hoffman does need to do much more work here if he is to avoid a solipsist conclusion to his view that user interface 'icons' hide reality.

4.4 Combination and Introspection

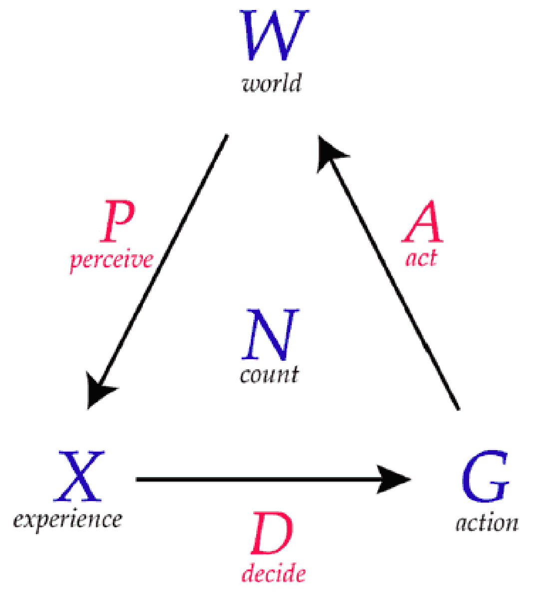

FIGURE 1 | A diagram of a conscious agent. A conscious agent has six components as illustrated here. The maps *P*, *D*, and *A* can be thought of as communication channels.

In this section, I will dig deeper into Hoffman and his collaborator's conceptual framework describing their proposed network of conscious agents. Hoffman and Prakash formalise their Conscious Realism theory by first defining a 'conscious agent' as consisting of three processes: perception, decision and action [Hoffman and Prakash 2014: 6]. (See their diagram reproduced here on the right.) It may seem that what Hoffman and Prakash mean by a 'conscious agent' here is a complete higher-order entity with consciousness, such as a dolphin or a human being. They write of how a 'conscious agent' 'chooses what actions to take based on the conscious experiences it has' and how it 'interacts with the world in light of the decision it has taken' [2014: 6]. Furthermore, in answering readers' qualms, they ascribe 'free will' to all conscious agents and goal-directed behaviour to some [2014: 14–15].

In spite of the teleological/intentionalist language they use to describe a 'conscious agent', what they mean by this is the atomic component of the world we live in. For example, they apply their model to visual perception in which each conscious agent represents one computational function of a visual system, working iteratively to build a visual percept [2014: 9–10]. For Hoffman and Prakash, then, each organism's visual system is comprised of multiple conscious agents working synchronously to form the organism's visual image.

In this respect, Hoffman and Prakash's theory draws on the atomistic elements of panpsychic micropsychism in which higher levels of conscious entities are built up by aggregating atomic components. How do they borrow from panpsychism? The debt is illustrated in this piece where Hoffman writes:

> When others see your face, they open a genuine, but limited, portal into your conscious world. Which is not to say that your face is conscious. It's not. You are conscious. Your face is an icon in the interface of the viewer.
>
> When I see a dog, my portal into consciousness is dimmer. I guess there is enjoyment of a bone and excitement by a squirrel. When I see an ant, my portal is dimmer still; I have little insight into the experiences behind my icon of an ant. With a rock, my portal is opaque; it offers no obvious insight into experiences behind the icon. My interface has, of necessity, finite limits; when it delivers a rock, it cries uncle — similarly, when it delivers atoms and molecules.

[Hoffman 2019b]

The difference between the two approaches is that panpsychic micropsychism allows for the reality of physical objects. Hoffman and Prakash's Conscious Realism may be cast as panpsychic micropsychism without matter. What the two ontologies do share, though, is the problem of combining perceptual experiences. How will Hoffman and Prakash combine perceptual experiences into what we see currently as meaningful bundles? How does the greenness, roundness and elasticity of a tennis ball all coalesce into the one perceptual 'object' and remain there. With a realist interpretation of our perceptual experiences, these combinations and regularities are explained easily and naturally as perceptual experiences are bundled with the individual physical object that causes in us those experiences. With Hoffman and Prakash's vast, interconnected network of atomic conscious agents, it's difficult to see a principled way in which to anchor particular combinations of experiences into stable bundles.

There is a second problem with combining experiences and one that they share with advocates of panpsychic micropsychism. As they put it:

> For instance, one's taste experiences of salt, garlic, onion, basil and tomato are somehow combined into the novel taste experience of a delicious pasta sauce. What is the relationship between one's experiences of the ingredients and one's experience of the sauce?
>
> [Hoffman and Prakash 2014: 11]

Hoffman and Prakash propose how their mathematical schemata could solve this combination problem through both creating Cartesian products and integrating over the space of perceptual experiences [2014: 12]. However, they can only formulate particular known combinations (such as complex tastes) in practice by borrowing from our current psycho-physical explanations. For example, we know that our experience of yellow can arise from combining our experiences of red and green. (This is how our television screens generate the thousands of different colours we see from only three primary colours.) We can describe how the separate wavelengths generated by a red object and a green object hit the photoreceptors in our eye to create the visual image of yellow. Our explanation of colour perception rests on our understanding of the properties of particular mind-independent objects and their interactions; namely the objects seen, the photons they emit or reflect and the cones in our retinas. In this way, Hoffman and Prakash's reverse engineering to arrive at their new equations using our current knowledge of physical interactions will be parasitic on realism.

Hoffman and Prakash realize that they will need to do better than this to convince the scientific community of their idealist scheme. They make an 'interesting prediction' suggested by their mathematical formalism; that the 'phenomenology of decision making is intimately connected with the spaces of perceptual experiences that are integrated in the decision process' [2014: 12]. We'll need to wait to see whether they are able to formulate a prediction with sufficient precision and then whether it will be confirmed.

A more serious problem they recognize is how to combine the individual subjects of experiences, the discrete conscious agents at the micro-level, into a unified conscious being at the macro-level, such as a human being or a dolphin. Here, Hoffman and Prakash develop two mathematical theorems from which they draw the conjecture that 'any subset of conscious agents from the pseudograph, adjacent to each other or not, can be combined to create a new conscious agent' [2014: 12].

The question immediately raised is: What is the basis for this conjecture? Why would we think that combining atomic conscious agents leads to a new conscious agent with its own distinct and 'particular phenomenological point of view' [2014: 12]. Why would we think this any moreso than a human family of individuals gives rise to an entirely new consciousness over and above the consciousnesses of the individuals who make up that family? In its intercommunications, a human family exhibits the interactions of perception, decision and action akin to that of Hoffman and Prakash's pseudograph of conscious agents. And yet, however we might formalise mathematically these intra-family interactions, there is nothing to suggest that a new supra-family consciousness pops out.

Hoffman and Prakash [2014: 12] refer to Coleman and Goff at some length disputing how such a view is even conceptually coherent given that

> ... a set of points of view have nothing to contribute *as such* to a single, unified successor point of view. Their essential property defines them against it: in so far as they are points of view they are experientially distinct and isolated—they have different streams of consciousness.
>
> [Coleman 2014]

Given both this paucity of demonstration of newness and the conceptual confusion over accessing private experiences, I suggest Hoffman and Prakash [2014: 12] are overstating their case significantly in thinking that their theorems 'give constructive proofs' of how atomic points of view can be combined to create a new point of view.

I suggest their overconfidence also extends to what they think their theorems say about introspection. They surmise that their method of mathematical combination shows how 'introspection emerges, in an intelligible fashion, from the combination of conscious agents' [2014: 13]. However, it's not at all clear how this introspection can occur as Hoffman and Prakash had already specified that a conscious agent has 'no direct experiential access to the sphere ... of experiences of any other conscious agent' [2014: 12]. Furthermore, in a later answer to an objector, they reinforce this inability of one conscious agent to experience directly the experience of another conscious agent. They write: 'The qualia X of a conscious agent C are private, in the sense that no other conscious agent C_i can directly experience X' [2014: 14]. Now, if we grant, for the sake of argument, the possibility of introspective access by one conscious agent into the private experiences of another, then that raises another difficulty: Why is it that I can't introspect the experiences of other conscious agents, such as that of my friend or my cousin or my dog?

The above considerations give rise to another similar quandary. And that is the question of how we should carve up aggregates of atomic conscious agents into distinct higher-order conscious agents, like you and me. If reality is a vast network of interconnected atomic conscious agents, then what determines the boundary between the aggregate of atomic conscious agents that constitute 'Leslie Allan' and the aggregates that separately constitute 'Donald Hoffman' and my dog? On a realist scheme, we easily and naturally draw the boundaries around conscious agents. I, Donald Hoffman and my dog have distinct physical bodies that really exist independently of looking at them. The challenge for Hoffman and Prakash is to carve up combinations of atomic conscious agents in a way that is not parasitic on the way we use our everyday realist scheme to individuate persons while also being principled.

Hoffman and Prakash's ascribing atomic conscious agents with the three processes of perception, decision and action gives the initial appearance that these conscious agents identify with us. These three processes also mirror the operation of neurons, with their inputs, internal processing and outputs. However, unlike physical organisms and neurons, Hoffman and Prakash's network of disembodied atomic conscious agents do not afford them the heuristic richness that comes with working with a physical substrate and an individuating principle that carves up micro-entities into natural wholes.

Consider also the limits of our communications with other conscious agents. Isn't it an extraordinary coincidence that the *only* collections of atomic conscious agents with which we are able to communicate are just those ones in which conscious awareness functions as the top-level modeller of the external world (the agent's environment) and its internal world (the agent's affective and conative states). Why is this a coincidence on Hoffman and Prakash's scheme? Because such top-level modelling is essential to our ability to collaborate with our tribal members in order for us to survive and reproduce. In other words, is it not a stroke of luck that the particular aggregates of atomic conscious agents that can communicate with each other are precisely the ones for which we can offer a realist scientific explanation for intra-species communication and the capacity for conscious awareness that underpins it?

4.5 Heuristic Sterility

In this section, I will elaborate more the criticism that Hoffman's idealism is missing the fundamental axioms and conceptual tools needed to develop his theory into an empirically rich paradigm that opens up new areas for enquiry and explanation. Take, for example, Hoffman's bold announcement that his reconstructions give 'mathematically precise theories about how certain conscious agents construct their physical worlds' [Hoffman 2008: 106]. Unfortunately for Hoffman, however, his reconstruction shows only how one conscious agent constructs their world. Despite his promise, his reformulation says nothing substantive about 'conscious agents and their dynamical interactions' [2008: 104]. His mathematical model is entirely silent on how, when I get a headache from a rock falling on my head when I'm all alone, this is the result of a 'decision' and 'action' of another conscious agent. His model says nothing about how, when I look up and see the moon, my moon experience is the result of the 'decision' and 'action' of other conscious agents in the network.

With the help of Prakash, Hoffman expended considerable effort in developing the mathematics of the supposed interactions between conscious agents. Drawing on the latest physics, they try to show that 'space-time and objects are among the symbols that conscious agents employ to represent the properties and interactions of conscious agents' [Hoffman and Prakash 2014: 13]. In working through sets of complex equations [2014: 13f], they claim to 'observe that the harmonic functions of the space-time chain that is associated with the dynamics of a system of conscious agents are identical to the wave function of a free particle; particles are vibrations not of strings but of interacting conscious agents' [2014: 13].

Hoffman and Prakash support this claim by identifying certain features of quantum theory with features of their conscious agents; features that they import explicitly into their theory. The mathematics employed is complex, so I'll leave it to others to evaluate their relevancy and veracity. In spite of all of this included complexity using eigenfunctions, Planck scale, Markovian dynamics, and so on, what is striking here is that Hoffman and Prakash afford us no insight into the most fundamental questions raised by their idealist mathematical formalism.

Again, their model gives us no inkling of how it is that we mentally construct stars and planets that, on our current understanding, existed for billions of years before any conscious entities appeared on the cosmic landscape. Or how it is that we conjure up forces of nature and particles so small that no conscious being can perceive directly.

Hoffman alludes to the difficult task he has set himself and that I am referring to here in his example of how neuroscientists stimulating a brain purport to bring about the experience of phosphenes.

> When, for instance, we stimulate primary visual cortex and see phosphenes, the cortex does not cause the phosphenes. Instead, certain interactions between conscious agents cause the phosphenes, and these interactions we represent, in greatly simplified icons, as electrodes stimulating brains.

[Hoffman 2008: 108]

It is exactly this story of how it is that direct 'dynamic interactions' between disembodied conscious agents cause the experience of phosphenes in the subject without the use of mind-independent physical electrodes that we are expecting from Hoffman and that we are missing entirely. With his tripartite model of the atomic consciousness, what we want to know is the mechanism behind a conscious agent's 'Decision'. How is this 'Decision' translated into an 'Action'? How does this 'Action' translate into a new 'Perception' had by another conscious agent? How is it that these conscious agents construct a multimode user interface (MUI) that is not just a little off the mark about the nature of reality, but is stupendously mistaken? Why do they construct a 'reality' in which physical entities and forces existed for billions of years even before the apparent first onset of conscious beings? Why do they construct a 'reality' in which it appears that motions of unconscious matter regularly determine the perceptual qualities, decisions and emotions of conscious entities? To what purpose is this universal self-deception?

Hoffman shows no appetite at all for exploring these fundamental questions, favouring instead a mathematical formalism that obscures more than it clarifies. He seems content to give away for free the remarkable progress gained over the last couple of centuries in our understanding of consciousness' neuronal base for a supposition that throws up many more mysteries than it purports to solve.

A solution to the mind-body problem will require bold ideas if we are to finally marry seamlessly the two domains of enquiry: the physical and the mental—just as Newton brought into the one conceptual framework terrestrial and celestial motions and Einstein combined the absolute categories of space and time into the one space-time dimension. In that respect, I welcome Hoffman's innovative conjecture. So far, however, Hoffman has given us only a very impoverished model of consciousness and of physical objects. In fact, we learn more about how human minds work from Plato and Aristotle writing more than two millennia ago. As far as becoming a fruitful and ongoing research programme, as we have seen, Hoffman's conjecture is yet to get some wind in its sails.

Let me pursue this problem of the heuristic sterility of Hoffman's Conscious Realism as it applies to the evolution of conscious agents in particular. Hoffman liberally uses biological explanations of adaption occurring in real time when doing so advances his thesis. For example, he writes:

> A backward retina, for instance, with photoreceptors hidden behind neurons and blood vessels, is not the "best" solution *simpliciter* to the problem of transducing light but, at a specific time in the phylogenetic path of *H. sapiens*, it might have been the best solution given the biological structures then available.
>
> [Hoffman 2009: 2]

However, Hoffman also informs us that his theory 'entails that DNA does not exist when it is not perceived' [Hoffman 2018: 11]. One problem for Hoffman here is that DNA has only been perceived within the last century. On Hoffman's scheme, with no DNA existing for the previous fourteen billion years or so, he has robbed himself of the mechanism of variation that underpins natural selection. Doing away with the DNA building blocks of evolution, that leaves Hoffman with no reason to hang on to the theory of evolution and no reason to generate mathematical simulations of evolutionary processes—except perhaps to satisfy his own intellectual curiosity.

Hoffman tries to meet this objection by pointing out that:

> Evolutionary changes in genes and body morphology can be modeled by evolution whether those genes and bodies are viewed as mind-dependent or mind-independent. The mathematics does not care. Nor does the fossil evidence. A dinosaur bone dated to the Jurassic can be interpreted along physicalist lines as a mind-independent object or, with equal ease, as a mind-dependent icon that we construct whenever we interact with a certain long-existing system of conscious agents.

[Hoffman 2008: 111]

Now, it may appear that Hoffman has up his sleeve an anti-realist interpretation of the evolution of organisms that is just as rich and fruitful as the scientific realist version. Not so. The 'dinosaur bone' he mentions here is simply a wished-for output of his mathematical model of how our visual systems present a bone. Hoffman and his colleagues have done none of the work on how a fossilized bone millions of years old is explained within the theoretical framework of a community of consciousness-only agents that arose only recently. It seems that crucial research program has not even begun.

Hoffman tells us:

> For the conscious realist there is, no doubt, interesting and fundamental work to be done here: We want a rigorous mathematical theory of the evolution of conscious agents which has the property that, when this evolution is projected onto the relevant MUIs, it gives us back the current physicalist model of evolution. That is, we must exhibit physicalist evolutionary models as special cases, in fact projections, of a richer and more comprehensive evolutionary theory.

[Hoffman 2008: 111]

How long will we need to wait for this richer and more comprehensive anti-realist evolutionary account of dinosaur bones and the birth of the universe? According to Hoffman, 'that's not going to be just a year or two. I mean, we're talking multi-decade effort here' [Tsakiris 2020]. In the meantime, in an interview with Frohlich, Hoffman does give us a hint of where he thinks evolutionary fitness might light lay:

> And there is a sense of "fitness", in vast social networks. The more connections you have, in some sense, the more fit you are. And the less connections you have, the less fit you are. So Google has tons of connections and the owners of Google are billionaires, Hoffman has a very few and he's not a billionaire.

[Frohlich 2019]

How heuristically fertile is Hoffman's idea? How will this sense of evolutionary 'fitness' in disembodied social networks explain the evolution of the eye 'icon' that appears to perceive things via light waves reflecting off the surfaces of mind-independent physical objects? How will that model of 'fitness' explain the evolution of the 'icon' for cosmic microwave background (CMB) radiation that exists from more than 13 billion years before conscious minds even arrived? How will that idea of 'fitness' explain the evolution of our belief that objects have permanence even when no one is perceiving them?

Paradoxically, Hoffman already relies on the standard scientific realist explanation of the characteristics of the human eye as resulting from selective environmental pressures acting over a long period of time. He writes:

> A backward retina, for instance, with photoreceptors hidden behind neurons and blood vessels, is not the "best" solution *simpliciter* to the problem of transducing light but, at a specific time in the phylogenetic path of *H. sapiens*, it might have been the best solution given the biological structures then available.

[Hoffman 2009: 2]

How will Hoffman's alternative explanation in terms of 'fitness' in disembodied social networks look when there is no time and space for evolution to work in and no photoreceptors, neurons and retina to work on?

Here again, Hofmann gives us a hope and prayer:

> Let's look at those dynamics and see which one makes sense. And then that will give us some insight into what's going on in the realm of conscious agents, what they're up to, and then when we project that back into our spacetime interface, we should get evolution by natural selection or hopefully a generalization that makes new predictions beyond evolution by natural selection.

[Frohlich 2019]

With just a promissory note for a research program that seems to not have even begun, it's difficult to see how 'fitness' seen as plentiful social connections will give Hoffman the conceptual resources he will need. By adopting this new heuristic, he seems especially handicapped. In abandoning our existing powerful heuristic that sees 'fitness' developing in a physical DNA substrate over immense periods of time and in wide geographical spaces, Hoffman seems to be tying both hands behind his back.

4.6 Post hoc Reconstruction

In the previous section, I examined how Hoffman's Conscious Realism appears to be wanting of the conceptual resources he will need to explain the evolution of species via backwards projection from his network of atomic conscious agents. The task he has set himself is even more onerous than this, for he hopes to explain not only evolution, but the successes of all branches of science, including quantum mechanics and general relativity [Frohlich 2019].

Unfortunately for Hoffman, any progress he thinks he has already made borrows from the realist framework he wants to ditch. Hoffman promises big: 'Conscious realism, by contrast, offers a scientific theory of the noumenal, viz., a mathematical formulation of conscious agents and their dynamical interactions' [Hoffman 2008: 104] and 'This notion can be made mathematically precise and yields experimental predictions' [2008: 105].

In this enterprise, Hoffman starts with vision science:

> We now have mathematically precise theories about how one type of conscious agent, namely human observers, might construct the visual shapes, colors, textures, and motions of objects (see, e.g., Hoffman 1998; Knill and Richards 1996, Palmer 1999).
>
> [Hoffman 2008: 106]

Hoffman explains how these mathematical formulae show how one conscious agent constructs in their visual field colour, texture, motion, depth, and so on. Note, however, that this work is neither unique to nor owes anything to Hoffman's Conscious Realism theory. Hoffman informs us that his mathematical models are simply reinterpretations of realist approaches to perception. He willingly concedes this point when he writes:

> Almost without exception the authors of these perceptual theories are physicalists who accept HFD and conceive of their theories as specifying methods by which human observers can *reconstruct* or approximate the true properties of physical objects that, they assume, exist objectively, i.e., independently of the observer (a claim about physical objects that is explicitly denied by conscious realism). But each of these perceptual theories can equally well be reinterpreted simply as specifying a method of object *construction*, not reconstruction.
>
> [Hoffman 2008: 106]

Here, Hoffman readily accepts that at the core of these physicalist theories is the axiom that there really are 'true properties' of mind-independent physical objects. But therein lay the reason for their historical predictive successes. It was in virtue of their proposing an invariant and known external cause of our perception that they were so successful. By contrast, the supposed predictive successes of Hoffman's mathematical models turn out to be not unique to Hoffman's conscious realism theory, but are parasitic on the successes of the realist research programs on human vision.

How easy is it for Hoffman to conscript the predictive successes of a realist view of perception as his own? Very easy. He tells us how:

> I can pull the *W* out of the model and stick a conscious agent in its place and get a circuit of conscious agents. In fact, you can have whole networks of arbitrary complexity. And that's the world.
>
> [Gefter 2016]

And working with Prakash, Hoffman does just that in a paper in which they simply redefine the space-time world of physics *W* to one in which the 'world *W* consists entirely of conscious agents' [Hoffman and Prakash 2014: 7].

Now, one mark of a progressive scientific theory is that it explains the predictive successes of its rival. For example, Einstein's Special Theory of Relativity explained the successful predictions of its rival, Newtonian dynamics. These Newtonian predictions included the return of Halley's Comet and the existence and location of the planet Neptune. Einsteinian Relativity would not have got a Guernsey if it had merely restated Newtonian mechanics in non-Euclidian geometry.

Hoffman and Prakash tap into this expectation when they write:

> The onus is on us to provide a mathematically rigorous theory of consciousness, to show how current physics falls out as a special case, and to make new testable predictions beyond those of current physics.
>
> [Hoffman and Prakash 2014: 15]

Here, Hoffman has hit the nail on the head. This is exactly what we need from his research programme if it is to gain some plausibility. What we are expecting are novel predictions derived from his theory that are later confirmed. And not simply *post hoc* mathematical reconstructions of pre-existing realist theories.

5. Conclusion

In this essay, I've examined in some detail the three components in Hoffman's ontology and philosophy of mind. I've devoted a section each to a critical examination of his Fitness Beats Truth (FBT) Theorem, Interface Theory of Perception (ITP) and, finally, his most radical idea, Conscious Realism.

Hoffman's Conscious Realism is a bold step in attempting to solve the mind-body problem and for that reason ought to be welcomed. However, as this critical review shows, it labours under some seemingly intractable problems. The problems that are solvable stem from linguistic muddles. Here, I have suggested alternative linguistic descriptions of how we should see 'icons' operating in a virtual reality world that remove these linguistic absurdities. The other problems are not so easy to overcome. Some writers had identified serious oversimplifications and misunderstandings with how Hoffman and his collaborators simulated evolution by natural selection.

I also articulated how Hoffman's thesis is self-defeating on a number of fronts. These included his reliance on evidences for biological evolution to disprove the truth of biological evolution, his dependence on the veridicality of our perception of distance, and his rejection of time and space when they are presumed by his evolutionary argument. Perhaps the most debilitating criticism of Hoffman's idealist theory is that it is heuristically sterile, lacking the theoretical resources required to solve the program's many puzzles. Coupled with this deficiency is the criticism that any claimed successes of Hoffman's program are simply *post hoc* reconstructions of the successes of scientific realism.

Hoffman's theory leaves many substantive questions unanswered with seemingly little interest from him to engage. An initial question is: What independent evidence does Hoffman offer for his network of atomic conscious agents? And what is his theory about the causal interconnections between these atomic constituents that give our sense-experience its regularity? There are other puzzles so far unaddressed. These include: What factors determine which atomic conscious agents will combine to form more complex entities? Why are the true connections and communications between conscious agents so opaque to the extent that what is presented to us is a gross mischaracterisation of reality at the most basic level? Why does this network of conscious agents appear positively deceitful?

Contrast these outstanding puzzles with our current models of reality. Our modern theories of cosmology, physics, biology and neuroscience combined tell a comprehensive and complex story that answers our questions about the nature of the physical universe and our place in it. In addition, over the span of more than two millennia, we have accumulated countless instances of confirmed novel predictions that attest to the veracity of our models of reality. Of course, the 'hard problem' of consciousness remains in philosophy of mind and physicists are still grappling with dark matter and dark energy. In spite of these unanswered questions, we have made enormous strides in our understanding of the universe and our place in it. Hoffman's theory, on the other hand, draws a big blank on these significant questions. In fact, it's just a promissory note for a theory that may or may not come later.

Granted, Hoffman was enticed into adopting his Conscious Realism schemata because, as he says, he was frustrated with the lack of progress on the mind-body problem.

As I have tried to show in this critical review, ditching scientific realism for Hoffman's version of idealism only serves to abandon a theoretical framework that has, over the last couple of centuries, proved enormously fruitful. The advances afforded us by scientific realism are not just in the theoretical domain, but also in the way it has led to improved medical and psychological treatments. With the insights gained by medical researchers, practitioners are now returning sight to the blind and movement to the paralysed. With our better understanding of brain function and mental illness, pharmacologists are changing the lives of patients suffering psychosis, depression, schizophrenia and a myriad of other psychological maladies. In our understanding of mind-independent matter, you would not be reading this on your electronic device if it were not for advances in quantum theory and we would not be flying one kilometre up in the air if it were not for advances in materials science. Giving up this scientific realist theoretical framework for a promise leaves us with no explanation for these stupendous successes.

In fairness to the newness of Hoffman's program, he and his collaborator recognize the immense challenges their theory faces. They concede:

> How can such an approach explain matter, the fundamental forces, the Big Bang, the genesis and structure of space-time, the laws of physics, evolution by natural election, and the many neural correlates of consciousness? These are non-trivial challenges that must be faced by the theory of conscious agents.

[Hoffman and Prakash 2014: 5]

However, they choose to cast these questions aside in favour of an abstract mathematical formalism: 'But for the moment we will postpone them and develop the theory of conscious agents itself' [2014: 5–6]. Considering the substantive challenges described in this critical review, I find it difficult to see how, even in principle, Hoffman's mathematical constructs will ever overcome these fundamental barriers. I would have liked to have seen Hoffman sketch out a conceptual framework that takes these problems seriously and that provides some heuristics for solving them, instead of focusing on developing impoverished mathematical models that shed no light on these fundamental problems.

Hoffman is right to feel frustrated at our lack of progress over some two millennia in solving the mind-body problem. That's not to say that no progress has been made. Advances in evolutionary psychology and biology, linguistics, cognitive science and information theory have greatly illuminated the internal workings of the mind and its development. Add to that advances that philosophers of mind have given us in developing the conceptual tools required for formulating and clarifying the problem. Part of the reason for the intractability of the mind-body problem is the sheer complexity of the brain and mind. The human brain is complex beyond all imagining, with an adult brain hosting some 80 billion neurons with some 100 trillion connections between them. The multiplicity of neurotransmitters shaping and modulating brain activity only adds to this unfathomable complexity.

What I see is that we are still in the pre-Newtonian phase of a final solution. We are waiting for a great unifier like Newton who, in a stroke of brilliance, brought into the one conceptual scheme what were considered in his time entirely disparate phenomena. Prior to Newton, natural philosophers were working with two sets of physics: one set for the terrestrial realm, guided by Buridan's dynamics, and another set for the celestial realm guided by Ptolemaic cosmology. Newton's Three Laws of Motion and Universal Law of Gravitation provided the glue that combined what seemed to be two entirely separate domains of enquiry. He forged a common language—a common set of equations and conceptual framework—that described the motions of both terrestrial and celestial bodies.

Is Hoffman the new Newton? Time will tell. However, this critical review should give us pause for unbridled optimism. Idealist theories have failed previously. Likewise, materialist theories have not gained universal acceptance. Both these approaches seek to reduce one domain to the other, relegating the secondary domain to that of a poor cousin, not really existing in its own right. Note how Newton's success was not bought at the cost of reducing either terrestrial phenomena or celestial phenomena to an expression of the other. I suspect that a solution to the mind-body problem will require us to give full credence to the primacy of each domain while affording us a novel conceptual scheme that explains the nature of both and their mutual interactions. Just as Newton provided us with his new physics.

I am grateful to Rached Blili and Vince Giuca for their many corrections to and comments on the pre-release version of this essay. I remain wholly responsible for any errors and omissions in the published version.

Initial draft release	Oct 25, 2020
First published	May 20, 2022

References

Abrahams, Marc 2012. Experiments Show We Quickly Adjust to Seeing Everything Upside-Down, *The Guardian*, URL = <https://www.theguardian.com/education/2012/nov/12/improbable-research-seeing-upside-down> (Published: 13 November, 2012).

Allan, Leslie 2016a. The Existence of Mind-Independent Physical Objects, URL = <https://www.rationalrealm.com/philosophy/metaphysics/mind-independent-physical-objects.html>.

Allan, Leslie 2016b. The Mind/Brain Identity Theory: A Critical Appraisal, URL = <https://www.rationalrealm.com/philosophy/metaphysics/mind-brain-identity-theory.html>.

Allan, Leslie 2016c. Towards an Objective Theory of Rationality, URL = <https://www.RationalRealm.com/philosophy/epistemology/objective-theory-rationality.html>.

Anderson, Chris 2019. The TED Interview: Donald Hoffman Has a Radical New Theory on How We Experience Reality, *TED Talks*, URL = <https://www.ted.com/talks/the_ted_interview_donald_hoffman_has_a_radical_new_theory_on_how_we_experience_reality> (Published: November, 2016).

Balkenius, C. and P. Gärdenfors 2016. Spaces in the Brain: From Neurons to Meanings, *Frontiers in Psychology* 7: 1820, doi: 10.3389/fpsyg.2016.01820 URL = <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5118439/> (Published: 22 November, 2016).

Berkeley, George 1710 (1974). Principles of Human Knowledge, in Berkeley's Philosophical Writings, ed. D. M. Armstrong, New York: Macmillan: 41–128.

Berman, Robby 2016. We Survive Because Reality May Be Nothing Like We Think It Is, *Big Think*, URL = <https://bigthink.com/robby-berman/we-survive-because-reality-may-be-nothing-like-we-think-it-is> (Published: 5 December, 2016).

Campbell, John O. 2016. Universal Darwinism as a Process of Bayesian Inference, *Frontiers in Systems Neuroscience* 10/49, doi: 10.3389/fnsys.2016.00049, URL = <https://www.frontiersin.org/articles/10.3389/fnsys.2016.00049/full> (Published: 7 June, 2016).

Cohen, J. 2015. Perceptual Representation, Veridicality, and the Interface Theory of Perception, *Psychonomic Bulletin and Review* 22: 1512–18, doi: 10.3758/s13423-014-0782-3, URL = <https://link.springer.com/article/10.3758/s13423-014-0782-3> (Published: 9 October, 2015).

Coleman, S. 2014. The Real Combination Problem: Panpsychism, Micro-subjects, and Emergence, *Erkenntnis* 79: 19–44, doi: 10.1007/s10670-013-9431-x.

Coyne, Jerry 2016. Does Evolution Lead Us to Perceive Reality, or Is It All an Illusion?, *Why Evolution Is True*, URL = <https://whyevolutionistrue.com/2016/07/21/does-evolution-lead-us-to-perceive-reality-or-is-it-all-an-illusion/> (Published: 21 July, 2016).

Dickinson, Kevin 2019. Did We Evolve To See Reality as It Exists? No, Says Cognitive Psychologist Donald Hoffman, *Big Think*, URL = <https://bigthink.com/surprising-science/does-reality-exist> (Published: 9 August, 2019).

Elaine, Howard Ecklund and Anne E. Lincoln 2011. Scientists Want More Children, *PLoS One* 6/8: e22590, doi: 10.1371/journal.pone.0022590, URL = <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3151251/> (Published: 5 August, 2011).

Else, Holly 2019. Nearly Half of US Female Scientists Leave Full-Time Science After First Child, *Nature*, URL = <https://www.nature.com/articles/d41586-019-00611-1> (Published: 19 February, 2019).

Frohlich, Joel 2019. Book Review and Author Interview: The Case Against Reality by Donald Hoffman, *Knowing Neurons*, URL = <https://knowingneurons.com/2019/12/06/book-review-case-against-reality/> (Published: 6 December, 2019).

Gefter, Amanda 2016. The Evolutionary Argument against Reality, *Quanta Magazine*, URL = <https://www.quantamagazine.org/the-evolutionary-argument-against-reality-20160421/> (Published: 21 April, 2016).

Goff, Philip, William Seager and Sean Allen-Hermanson 2015. Panpsychism, *The Stanford Encyclopedia of Philosophy* (Summer 2020 Edition), ed. Edward N. Zalta, URL = <https://plato.stanford.edu/archives/sum2020/entries/panpsychism/>.

Guyer, Paul and Rolf-Peter Horstmann 2015. Idealism, *The Stanford Encyclopedia of Philosophy* (Winter 2019 Edition), ed. Edward N. Zalta, URL = <https://plato.stanford.edu/archives/win2019/entries/idealism/>.

Hoffman, Donald D. 2008. Conscious Realism and the Mind-Body Problem, *Mind and Matter* 6/1: 87–121, URL = <http://www.cogsci.uci.edu/~ddhoff/ConsciousRealism2.pdf>.

Hoffman, Donald D. 2009. The Interface Theory of Perception: Natural Selection Drives True Perception to Swift Extinction, URL = <http://www.cogsci.uci.edu/~ddhoff/interface.pdf>.

Hoffman, Donald D. 2018. The Interface Theory of Perception, *The Stevens' Handbook of Experimental Psychology and Cognitive Neuroscience Volume 2*: 1-24, doi: 10.1002/9781119170174.epcn216, URL = <http://cogsci.uci.edu/~ddhoff/Hoffman-Stevens-Handbook.pdf>.

Hoffman, Donald D. 2019a. Is Reality Real? How Evolution Blinds Us to the Truth about the World, *New Scientist*, URL = <https://www.newscientist.com/article/mg24332410-300-is-reality-real-how-evolution-blinds-us-to-the-truth-about-the-world/> (Published: 31 July, 2019).

Hoffman, Donald D. 2019b. Imitation and Extinction: The Case Against Reality, *Los Angeles Review of Books Quarterly Journal: Imitation, No. 23*, URL = <https://lareviewofbooks.org/article/imitation-extinction-case-reality/> (Published: 18 August, 2019).

Hoffman, Donald D. and Chetan Prakash 2014. Objects of Consciousness, *Frontiers of Psychology* 5/577, doi: 10.3389/fpsyg.2014.00577, URL = <https://www.frontiersin.org/articles/10.3389/fpsyg.2014.00577/full> (Published: 17 June, 2014).

Hoffman, Donald D., Manish Singh and Chetan Prakash 2015a. The Interface Theory of Perception, *Psychonomic Bulletin and Review* 22/6: 1480-506, doi: 10.3758/s13423-015-0890-8, URL = <https://link.springer.com/content/pdf/10.3758/s13423-015-0890-8.pdf> (Published: 18 September, 2015).

Hoffman, Donald D., Manish Singh and Chetan Prakash 2015b. Probing the Interface Theory of Perception: Reply to Commentaries, *Psychonomic Bulletin and Review* 22: 1551–76, doi: 10.3758/s13423-015-0931-3, URL = <https://link.springer.com/content/pdf/10.3758/s13423-015-0931-3.pdf>.

Kappers, A. M. L. 1999. Large Systematic Deviations in the Haptic Perception of Parallelity, *Perception* 28: 1001–12, URL = <https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.294.8163&rep=rep1&type=pdf>.

Koenderink, J. J. 2014. The All Seeing Eye? *Perception* 43: 1–6, URL = <https://journals.sagepub.com/doi/pdf/10.1068/p4301ed>.

Koenderink, J. J., A. J. van Doorn, H. de Ridder and S. Oomes 2010. Visual Rays Are Parallel, *Perception* 39: 1163–71, URL = <https://journals.sagepub.com/doi/pdf/10.1068/p6530>.

Leiter, Brian 2019a. Evolution Does Not Select for Veridical Perception, *Leiter Reports: A Philosophy Blog*, URL = <https://leiterreports.typepad.com/blog/2019/09/evolution-does-not-select-for-veridical-perception.html> (Published: 17 September, 2019).

Leiter, Brian 2019b. More on Whether Evolution Really Selects for Veridical Perception, *Leiter Reports: A Philosophy Blog*, URL = <https://leiterreports.typepad.com/blog/2019/10/more-on-whether-evolution-really-selects-for-veridical-perception.html> (Published: 3 October, 2019).

Martin, Brian 2019. Do We See Icons or Reality? A Review of Donald Hoffman's The Case Against Reality, *Social Epistemology Review and Reply Collective* 8/12: 13–24, URL = <https://social-epistemology.com/2019/12/05/do-we-see-icons-or-reality-a-review-of-donald-hoffmans-the-case-against-reality-brian-martin/> (Published: 5 December, 2019).

Martínez, Manolo 2019. Usefulness Drives Representations to Truth: A Family of Counterexamples to Hoffman's Interface Theory of Perception, *Grazer Philosophische Studien* 96: 319-41, doi: 10.1163/18756735-09603004, URL = <http://philsci-archive.pitt.edu/15846/1/article.pdf>.

Mausfeld, Rainer 2015. Notions Such as 'Truth' or 'Correspondence to the Objective World' Play No Role in Explanatory Accounts of Perception, *Psychonomic Bulletin and Review* 22/6: 1535–40, doi: 10.3758/s13423-014-0763-6, URL = <https://www.researchgate.net/publication/282045248>.

McDermott, Amy 2020. Neurons in the Hippocampus Can Make Several Different Maps of the Same Environment, *Journal Club*, URL = <https://blog.pnas.org/2020/04/neurons-in-the-hippocampus-can-make-several-different-maps-of-the-same-environment/> (Published: 17 April, 2020).

Metzinger, T. 2003. *Being No One: The Self-model Theory of Subjectivity*, Cambridge, MA: MIT Press.

Moser, E. I., E. Kropff and M. B. Moser 2008. Place Cells, Grid Cells, and the Brain's Spatial Representation System, *Annual Review of Neuroscience* 31: 69–89, doi: 10.1146/annurev.neuro.31.061307.090723.

Murphy, Paul Austin 2020a. Donald Hoffman's Eye Candy: Conscious Realism's Mathematical Models, *Paul Austin Murphy's Philosophy*, URL = <http://paulaustinmurphypam.blogspot.com/2020/01/donald-hoffmans-eye-candy-conscious.html> (Published: 29 January, 2020).

Murphy, Paul Austin 2020b. Donald Hoffman's Long Jump from Evolutionary Biology/Theory to Highly-Speculative Philosophy, *Medium*, URL = <https://medium.com/@paulaustinmurphy2000/donald-hoffmans-long-jump-from-evolutionary-biology-theory-to-highly-speculative-philosophy-c12a09490ba2> (Published: 6 February, 2020).

Murphy, Paul Austin 2020c. Donald Hoffman's Conscious Realism vs. Panpsychism and Idealism, *Medium*, URL = <https://medium.com/@paulaustinmurphy2000/donald-hoffmans-conscious-realism-vs-panpsychism-and-idealism-412dceaac6f0> (Published: 13 February, 2020).

Ogrady, Jane 2019. Donald Hoffman's The Case Against Reality Is Hard to Get Your Head Around, *The Spectator*, URL = <https://www.spectator.co.uk/article/donald-hoffman-s-the-case-against-reality-is-hard-to-get-your-head-around> (Published: 12 October, 2019).

Pont, S. C., H. T. Nefs, A. J. van Doorn, M. W. A. Wijntjes, S. F. te Pas, H. deRidder and J. J. Koenderink 2012. Depth in Box Spaces. *Seeing and Perceiving* 25: 339–49, URL = <http://www.maartenwijntjes.nl/oldy/site/Publications_files/boxspaces.pdf>.

Prakasha, Chetan, Kyle D. Stephens, Donald D. Hoffman, Manish Singh and Chris Fields 2017. Fitness Beats Truth in the Evolution of Perception, Under review, URL = <http://cogsci.uci.edu/~ddhoff/FBT-7-30-17> (Retrieved: 2 July, 2020).

Reynolds. Hugh D. 2019. Reality Check, *Physics World* (August 2019 edition), URL = <https://physicsworld.com/a/reality-check/>.

Roser, Max 2017. Fertility Rate, *Our World in Data*, URL = <https://ourworldindata.org/fertility-rate> (Revised: 2 December, 2017).

Shermer, Michael 2015. Did Humans Evolve to See Things as They Really Are?, *Scientific American*, URL = <https://www.scientificamerican.com/article/did-humans-evolve-to-see-things-as-they-really-are/> (Published: 1 November, 2015).

Tsakiris, Alex 2020. Dr. Donald Hoffman, Materialism's Final Death Blow? |436|, *Skeptiko*, URL = <https://skeptiko.com/dr-donald-hoffman-materialisms-final-death-blow-436/> (Published: 7 January, 2020).

Vlerick, Michael 2014. Natural Selection Does Care about Truth, *International Studies in the Philosophy of Science* 28/1: 65-77, doi: 10.1080/02698595.2014.915651, URL = <https://philpapers.org/rec/BOUNSD-2>.

Wilson, Andrew D. 2015. The Interface Theory of Perception - The View from Ecological Psychology, *Notes from Two Scientific Psychologists*, URL = <http://psychsciencenotes.blogspot.com/2015/09/the-interface-theory-of-perception-view.html> (Published: 28 September, 2015).

