The Simulation Hypothesis and Meta-Problem of Everything

Marcus Arvan
University of Tampa
marvan@ut.edu

Abstract: In a new paper, David. J. Chalmers examines eleven possible solutions to the *meta-problem of consciousness*, 'the problem of explaining why we think that there is a problem of consciousness.' The present paper argues that Chalmers overlooks an explanation that he has otherwise taken seriously, and which a number of philosophers, physicists, and computer scientists have taken seriously as well: the hypothesis that we are living in a *computer simulation*. This paper argues that a particular version of the simulation hypothesis is at least as good of a solution to the meta-problem of consciousness as many explanations Chalmers considers, and may even be a better one—as it may be the best solution to a much broader meta-philosophical problem: the *meta-problem of everything*, the problem of explaining why our world has the quantum-mechanical, relativistic, and philosophical features it does.

In a new paper, David. J. Chalmers examines the *meta-problem of consciousness*, 'the problem of explaining why we think that there is a problem of consciousness.'[1] According to Chalmers, this is an empirical problem—one concerning the mechanisms that lead people to believe and report that there is a hard problem of consciousness.[2] Chalmers then examines eleven possible explanations of the meta-problem, that is, eleven empirical hypotheses about why people *believe and say* that there is a hard problem of consciousness. The present paper argues that Chalmers overlooks a solution to the meta-problem which he has otherwise taken seriously[3], and which a number of philosophers[4], physicists[5], computer scientists[6], and programmers[7] have taken seriously as well: the theory that we live in a *computer simulation*. It is not altogether surprising that Chalmers ignores this hypothesis in discussing the meta-problem of consciousness—as the simulation hypothesis has been argued by some to lack

[1] Chamers (2018a): 6.
[2] Ibid: 10.
[3] Chalmers (2015, 2017, 2018b).
[4] Bostrom (2003), Arvan (2013, 2014, 2015), Johnson (2011), Mizrahi (2017).
[5] Beane et al (2012) and Campbell *et al*. (2017).
[6] Moravec (1998), and Whitworth (unpublished manuscript).
[7] Grange (2016).

any evidential support.[8] However, this paper argues to the contrary that a particular version of the simulation hypothesis has at least as much evidential support as many of the hypotheses Chalmers considers, while also possessing far greater explanatory power—as the simulation hypothesis in question may be best solution to a much broader metaphilosophical problem: the 'meta-problem of everything.'

§1. The Case for the P2P Simulation Hypothesis

Our world has a wide variety of deeply perplexing physical and philosophical features. Consider physics. At present, our two best theories of fundamental physics are the General Theory of Relativity, which explains gravitation, and Quantum Mechanics, which explains all other known forces. Both theories have been systematically confirmed by experiment—yet both theories tell us our world's physics is incredibly strange. General Relativity tells us that:

i. ***Space and time are relative to observers***: simultaneous events in one reference frame are non-simultaneous from another, time moves at different rates depending on the observer's frame of reference, and the physical properties of objects in space-time (e.g. their length) depends on the observer's reference-frame.

ii. ***The physical world has a 'speed-limit':*** no information can travel faster than light.

Quantum mechanics, in turn, tells us that all of the following are true of our world:

iii. ***Quantum superposition:*** every particle simultaneously exists in many different *eigenstates* (i.e. a superposition of different space-time locations and properties).

iv. ***Quantum indeterminacy:*** the eigenvalue a particle will be observed to have upon measurement is indeterminate, in that the value can *in principle* only be predicted probabilistically.

[8] Huemer (2016).

v. ***Wave-particle duality***: every individual particle simultaneously has properties of particle (existing at a particular point) and a wave (spread out over space and time).

vi. ***Wave-function collapse:*** observation of a particle (or measurement of quantum system it is a part of) leads the wave-like features of a particle (viz. the particle's superposition) to 'collapse' to a single observed value (i.e. the observed properties of the particle).

vii. ***Quantum entanglement***: particles arbitrary distances apart can become *entangled*, such that changing the physical properties of one particle will *instantaneously* change the other particle's properties without any observable exchange of information.

viii. ***Minimum space-time distance:*** there is a minimum space-time distance below which space and time themselves have no physical meaning (the Planck Length).[9]

ix. ***Quantum retrocausality:*** measurements of a quantum system can have observable effects on the system *earlier in time*, causing wave-function collapse *before* the measurement is taken.[10]

These features of our world are incredibly bizarre—yet they are implied by the equations of quantum mechanics, and quantum mechanics been systematically confirmed by experiment.

Finally, our world has a number of puzzling *meta*physical features—among them (to simplify a great deal):

x. ***The mind-body problem (hard problem of consciousness):*** it appears impossible to reduce or identify phenomenal consciousness to anything physical.[11]

[9] Padmanabhan (1985).
[10] Leifer & Pusey (2017).
[11] See e.g. Chalmers (2016).

xi. ***The problem of causation:*** events in our world are 'constantly connected' (viz. causal regularities), yet empirical observation appears insufficient to explain *why* events are constantly connected, or whether there is a primitive metaphysical 'causal force' that connects them.[12]

xii. ***The problem of time and time's passage:*** although some arguments in physics and philosophy suggest that all times (past, present, and future) exist eternally, time seems to *pass*.[13]

xiii. ***The problem of personal identity:*** although we experience ourselves as though we persist as identical persons across time, it appears impossible to explain personal identity in physical or psychological terms.[14]

xiv. ***The problem of free will:*** philosophical considerations and the laws of physics suggest that our choices must be determined—which some philosophers think entails we have no free will. Yet it seems, for all that, like we *have* free will.[15]

All of these puzzling features of our world—the puzzling features of physics, and the philosophical problems just presented—are typically grappled with *independently*: with the physicists doing physics, and philosophers doing the philosophy. Both groups, however, have run up against apparently insuperable obstacles. On the one hand, physicists have been unable to explain *why* our world has relativistic and quantum-mechanical features. Physics merely studies how the world actually behaves, basing its equations and explanations on observation. However, that is all physics can do: explain *what* we observe and how what we

[12] See Schaffer (2016) for an overview.
[13] Markosian (2016), Le Poidevin (2015).
[14] Ninan (2009).
[15] O'Connor & Franklin (2018).

observe ‘works' (viz. the equations of relativity and quantum theory). What physics cannot do is explain *why* our world has the observable features in the first place. Why, of all of the metaphysically possible universes that could have existed, do we exist in world with relativistic and quantum mechanical physics? This does not appear to be a question of physics, but rather of *metaphysics*. Alas, metaphysics arguably faces insuperable problems of its own. Metaphysical debates on most major issues—ranging from mind-body problem to problems of time, personal identity, free will, and so on—typically result in interminable stalemates: with different groups of philosophers defending different, fundamentally opposed metaphysical theories, with no clear way to resolve which theory is true.[16] For example, in the mind-body problem literature alone, there are serious proponents of eliminative materialism, mind-brain identity-theory, non-reductive physicalism, functionalism, panpsychism, property dualism, and substance dualism.[17] Similar stalemates exist across metaphysics, and in philosophy in general—leading some to wonder whether philosophy makes real progress.[18] Another way of putting this is that physics and philosophy together face what we might call **the meta-problem of everything**: the question of explaining why our world spears to us to have the many puzzling physical and metaphysical features it does, including the meta-problem of consciousness. Might there be a single, unified solution to this broader meta-problem that in turn solves the meta-problem of consciousness? Indeed, there may be.

[16] See Willard (2013).
[17] Bogardus (2013).
[18] Dietrich (2011), Slezak (2018).

Consider the hypothesis that we are living in a computer simulation. Do we have any *evidence* for the simulation hypothesis beyond Bostrom's probabilistic speculation[19]? In a recent article, Arvan notes that although each of the following theories is controversial, philosophers and physicists have argued there is some evidence for each of them[20]:

A. ***Eternalism:*** past, present, and future objects and properties all exist "timelessly."
B. ***The Multiverse Hypothesis:*** the observable universe is a small part of a vast 'multiverse.'
C. ***The Holographic Principle:*** to unify quantum mechanics and general relativity, the universe must be understood as digital information written on the cosmological horizon.
D. ***Mind-body Dualism:*** the mind is in some sense non-physical.
E. ***Subjectivity About the Flow of Time:*** time's passage is not in the objective physical world but rather within us (i.e. within consciousness).
F. ***The Further Fact Theory of Personal Identity*****:** personal identity is a brute, simple fact that cannot be reduced to any physical or psychological phenomena.
G. ***Single Commonly-Experienced (or"Actualized") Timeline:*** only one physical universe—our Universe—is experienced by conscious observers.

Arvan argues that if all of these theories are jointly true, then our 'physical world' is a *hologram* generated by each individual's *consciousness* 'reading' digital information on the cosmological horizon, *projecting* that digital information as a four-dimension 'world' of

[19] Bostrom (2003).
[20] Arvan (2013): Section I.

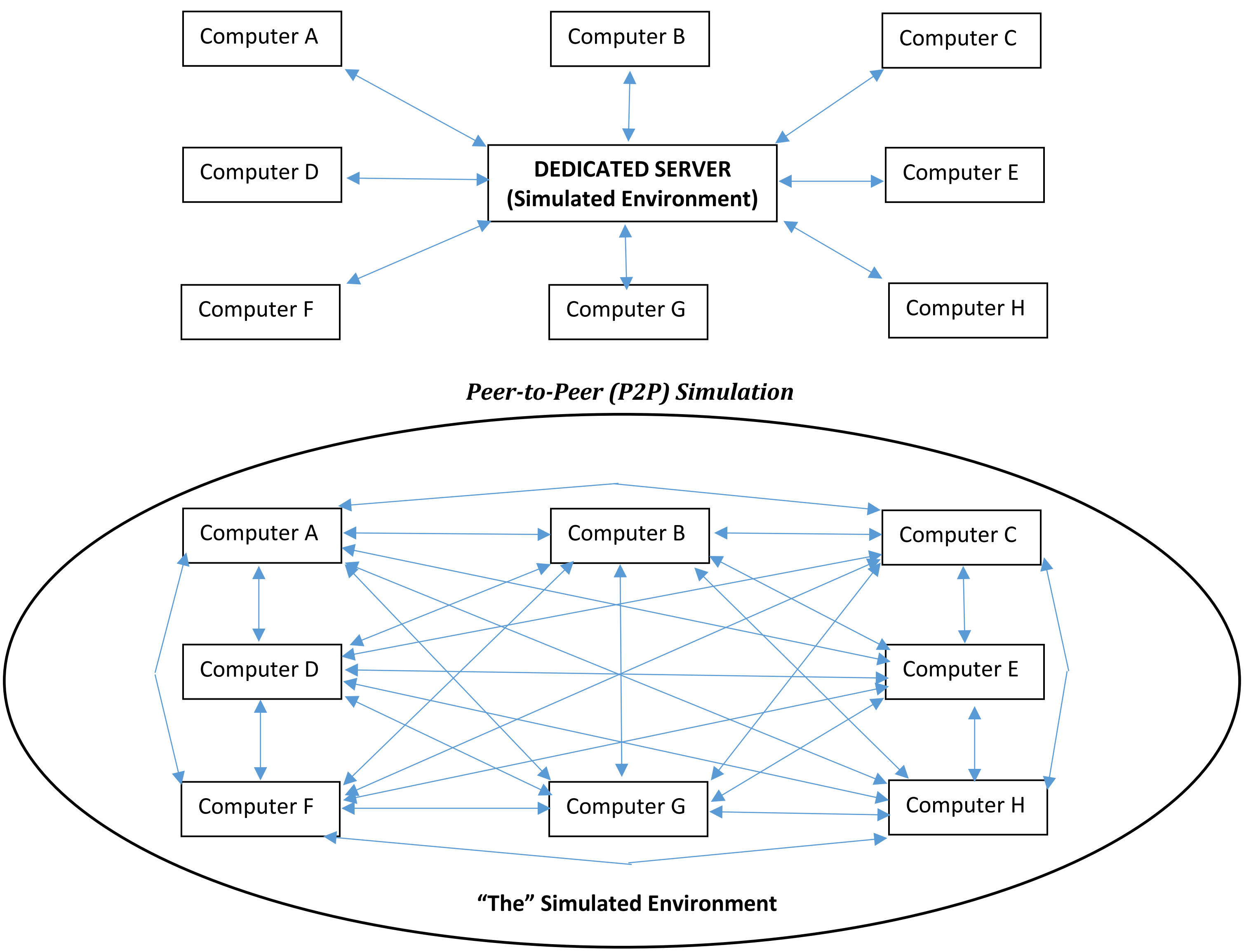
objects in space-time, such that the *joint projections* of each person's consciousness constitute an *intersubjective reality* that we all experience together.[21]

In the same article and in several follow-up pieces, Arvan argues that this metaphysical model of reality is functionally identical to a certain kind of computer simulation: a *peer-to-peer networked* (P2P) simulation.[22] First, Arvan notes that online videogames *just are* physical mechanisms (computer processor) *reading* digital information (e.g. on a DVD), *projecting* that information as a four-dimensional world for different 'users' to navigate—which is what hypotheses (A)-(G) jointly entail our world is. Second, Arvan argues that a particular kind of simulation—**peer-to-peer networked (P2P) simulations**—actually replicate our world's relativistic and quantum-mechanical physical features due to the computational structure of peer-to-peer networking itself.[23] For consider what a P2P simulation is. In contrast to dedicated server simulations—where there is a central computer representing the spatio-temporal locations of all objects in the simulation—a P2P simulation has no central computer at all: instead, a P2P simulation is simply a network of independent simulations interacting with each other (see Figure 1). In a P2P simulation, each 'user' only ever experiences *their* simulation, and 'the physical world' that all users experience in common is *just a superposition of all of the simulations interacting on the network*.

[21] Ibid: Section II.
[22] Ibid: footnotes 60-61; Arvan (2014, 2015).
[23] Arvan (2014, 2015).

Figure 1. Two Types of Simulations

Dedicated Server Simulation



Next, Arvan shows[24] that because there is no central server representing where objects 'really are' in a P2P simulation, P2P simulations computationally replicate every basic feature of quantum mechanics:

[24] Arvan (2014, 2015).

- ***Replicating quantum superposition:*** A P2P simulation just is a *superposition* of multiple simulated environments interacting in parallel.
- ***Replicating quantum indeterminacy:*** 'The' location of any object or property in a P2P simulation is therefore *indeterminate*, given that each computer on the network has its own representation of where 'the' object or property is, with no dedicated server on the network to represent where the object or property 'really' is.
- ***Replicating wave-particle duality:*** Because different simulations in a P2P network represent the same objects in slightly different positions at any given instant, a dynamical description of where a given object/property probably is in 'the environment' will have *features of a wave* (viz. an amplitude equivalent to the number of computers representing the object at a given point, and wavelength equivalent to dynamical change of how many computers represent the object at a given point at the next instant). At the same time, because each individual simulation always has its *own* representation of where objects are in the environment, any measurement taken by any simulation in the network will always represent fundamental objects as existing at a particular point (*qua* particle).
- ***Replicating wave-function collapse*:** Because a P2P simulation is superposition of parallel interacting simulations, but any measurement by a user *in any one simulation* will represent that object at a determinate location *within their individual simulation*, any measurement taken in a P2P simulation will appear to any observer to 'collapse' the superposition/wave-like properties of any object to a *point*.

- ***Replicating quantum entanglement:*** If a P2P simulation does not have perfect error-correction, then multiple computers on a network can represent one particle as existing in two places, representing them as *one entangled particle*.
- ***Replicating minimum space-time distance***: Because a simulated reality is comprised by digital rather than continuous information, that reality must be 'pixelated' at a fundamental level—that is, its smallest objects and properties must be *separated* by some minimum distance ('between the digital information') that *has no informational content* in the simulation.
- ***Replicating quantum retrocausality:*** in P2P simulations, spatiotemporal conflicts between different simulations on the network may be resolved by error-correction algorithms that 'alter the past.'[25]

The P2P simulation not only replicates quantum features—it replicates relativistic ones:

- ***Replicating spatio-temporal relativity:*** P2P simulations have no 'master clock' or objective representation where things are in their simulated space-time: *all* observations are relative to individual users on each individual simulation. Further, Grange argues that processing limitations in a P2P system should have relativistic effects on observed time and space.[26]
- ***Replicating maximum speed-limit:*** Grange argues that in a P2P simulation, bandwidth limitations constraining communication-speed between different

[25] Here is one actual example: in the videogame Halo 3, if I kill your character slightly before you kill me on my simulation, but the opposite occurs on *your* simulation (i.e. you kill me slightly before I kill you), the network retroactively resolves the temporal conflict by *killing both characters* 'simultaneously' (something which notoriously frustrated many gamers).

[26] Grange (2016).

simulations on the network entail a *maximum speed limit* of information travel within each simulation.[27]

Finally, Arvan shows that P2P simulations replicate a variety of metaphysical problems:

- ***Replicating the mind-body problem:*** Because each individual's 'subjective point of view' is constituted by a *processor* that underlies and generates their 'physical' reality as a *projection*, each individual in a P2P simulation would have the sense that their 'mind' cannot be reduced to anything physical-functional in their world…**and they would be right** (the processor they *are* is not in the simulation at all: it *grounds* the simulation as the *projecting mechanism* that represents 'their world' as a hologram).
- ***Replicating the problem of causation:*** because each individual in a P2P system would experience their reality as a connected series of events, each individual would have the sense that there must *be* something to causation beyond the series events itself—some 'force' that explains why those events are connected…**and they would be right** (their *processor* connects the events).
- ***Replicating the problem of time, time's passage, and subjectivity of time's flow:*** Because a P2P simulation consists of digital information 'being read' by a processing mechanism, individuals living in a P2P simulation would believe there is a sense in which past, present, and future exist 'timelessly', while also believing time passes subjectively…**and they would be right** (as time's passage would consist in their processor *processing* 'timeless' digital information, *projecting* a 'moving present').
- ***Replicating problems of personal identity and free will:*** Individuals living in a P2P simulation would be inclined to believe that there is something more to their personal

[27] Ibid.

identity and free will beyond their 'physical world'…**and they would be right** (their identity over time would be comprised by the *processor* that grounds their projected point-of-view, and their 'choices' in the simulation would actually be made in a *higher-reference frame* (the level of the 'user' or processor'), giving rise to *apparent causal-closure* in the simulation (and hence, the worry that they are not free).[28]

The P2P Simulation Hypothesis is, to my knowledge, the only unified explanation currently on offer of all of the above physical and meta-physical features of our world: that is, it is the only unified explanation we have of the **meta-problem of everything—**of which the meta-problem of consciousness is a special case.

2. Comparison to Alterative Solutions to the Meta-Problem of Consciousness

Chalmers considers the following eleven explanations of the hard problem of consciousness (each of which thus constitutes a 'solution' to the meta-problem of consciousness):

1. *Introspective models*: by modeling its own internal states, the brain represents its own states in a way that gives rise to the hard-problem of consciousness.[29]
2. *Phenomenal concepts*: the hard-problem results from special concepts presenting our conscious states to us as otherwise than physical.[30]
3. *Independent roles*: the hard-problem results from phenomenal and physical concepts lacking strong inferential connections to one another.[31]

---

[28] As a simple example, consider the videogame PacMan. From *within* PacMan's simulated world, just about everything appears 'deterministic'. Even *PacMan's* behavior—the character you control as a user—can be predicted probabilistically, given your tendencies as a player. However, PacMan's behavior (unbenknownst to anyone 'living in' that reality) is *actually* controlled by you as the outside user. PacMan's behavior in the simulation, then, is *not actually determined* by the 'physical laws' of the simulation (including whatever probabilistic equations explain how he behaves). His behavior just *appears* to be determined because no one in the simulation has any observational access to the *inputs* to the system you are making as its outside 'user.'
[29] Chalmers (2018a): 12-3.
[30] Ibid: 13-4.
[31] Ibid: 14.

4. *Introspective opacity*: the hard-problem results from brain-processes representing other brain-processes as though they are *not* brain processes.[32]
5. *Direct access*: the hard-problem results from introspective states being direct and non-inferential, representing things like 'greenness' as primitive properties.[33]
6. *Primitive quality attribution*: the hard-problem results from our perceptual capacities *attributing* primitive qualities (e.g. colors) to things.[34]
7. *Primitive relation attribution*: the hard-problem results from introspective models introducing primitive relations of seeing, hearing, etc., to simplify highly complex relations (such as color-wavelengths) in a cognitively efficient way.[35]
8. *Introjection and the phenomenological fallacy*: the hard-problem results from fallaciously inferring from phenomenal experiences (e.g. consciousness experience of redness) that the *object* of the experience (phenomenal redness) *exists*.[36]
9. *The user illusion*: the hard-problem is the result of consciousness itself being an *illusion* generated by the brain, much like folders on computer desktop are illusions regarding the computational reality that underlies them.[37]
10. *The use-mention fallacy*: the hard-problem results from the way we think about consciousness being different than the way we think about brain states.[38]

[32] Ibid: 14-5.
[33] Ibid: 15-6.
[34] Ibid: 16-8.
[35] Ibid: 18-20.
[36] Ibid: 20-1.
[37] Ibid: 21.
[38] Ibid: 21-2

11. *Historical explanations*: the hard-problem can be explained away in either evolutionary terms[39], re-entrant feedback loops in a higher-dimensional space[40], conflicts in judgments about consciousness from dual-process cognitive systems[41], the necessity of cognitive system to avoid a regress in positing subject-object distinctions[42], and so on.[43]

How should we evaluate each of these explanations against each other, and against the P2P Simulation Hypothesis? Although there are complex issues in the philosophy of science here, two criteria immediately present themselves: **empirical adequacy** (support from empirical evidence) and **explanatory power**. For example, whereas Ptolemaic astronomy falsely predicted planetary orbits, and could not explain observations of retrograde motion, Copernican astronomy explained and predicted retrograde motion. Finally, empirical adequacy can be broken down into roughly two issues: how much *positive empirical support* a hypothesis has (viz. confirmation), and the extent to which the hypothesis *conflicts* with observations (viz. disconfirmation).

Let us consider then, first, how much positive empirical support Chalmers' eleven hypotheses have. To the best of my knowledge, all eleven hypotheses are based primarily on *philosophical conjecture*: I do not know of a single empirical study that suggests the existence of phenomenal concepts, introspective opacity, and so on. Because the P2P Hypothesis is largely based on conjecture as well (viz. seven philosophical and physical hypotheses philosophers and physicists have argued there is some evidence for), the P2P Hypothesis

[39] Humphrey (2014),
[40] Ibid.
[41] Fiala et al (2011).
[42] Molyneux (2012).
[43] Chalmers (forthcoming): 22-3.

appears to have roughly the *same level* of positive empirical support as many of the eleven hypotheses Chalmers considers.

Now consider disconfirmation: the question how much different hypotheses *conflict* with our observed evidence. The most obvious concern to have here about the P2P Hypothesis is that, insofar as it explains the problem of consciousness in functionalist terms (viz. simulated reality), it cannot explain *phenomenal properties* themselves (i.e. 'what it is like' to phenomenally experience the color red). However, as Chalmers argues, this appears to be a problem with *many* of the hypotheses he considers.[44] Thus, the P2P Hypothesis fares *no worse* than many of the hypotheses that Chalmers considers as serious explanations of the meta-problem of consciousness.

Finally, what about explanatory power? Here, the P2P Hypothesis is a clear winner. For whereas the eleven hypotheses Chalmers considers would at best explain why we think there is a problem of consciousness, the P2P Hypotheses may provide the first unified explanation of a much wider variety of physical and philosophical problems—not just the **meta-problem of consciousness**, but also the **meta-problem of everything**.

In sum, taking into account empirical adequacy and explanatory power together, the P2P Hypothesis is at least as good of a solution to the meta-problem of consciousness as many of the hypotheses Chalmers considers—and it may well be the best explanation of all.

[44] See e.g. ibid: 21, 22, 25-7, 30-3.

References

Arvan, Marcus (2015). The Peer-to-Peer Simulation Hypothesis and a New Theory of Free Will. *Scientia Salon*.

----- (2014). A Unified Explanation of Quantum Phenomena? The Case for the Peer-to-Peer Simulation Hypothesis as an Interdisciplinary Research Program. *Philosophical Forum* 45 (4):433-446.

-----(2013). A New Theory of Free Will. *Philosophical Forum* 44 (1):1-48.

Beane, S. R., Davoudi, Z., & Savage, M. J. (2012). Constraints on the Universe as a Numerical Simulation. *arXiv preprint arXiv:1210.1847*.

Bogardus, Tomas (2013). Undefeated dualism. *Philosophical Studies* 165 (2):445-466.

Bostrom, Nick (2003). Are we living in a computer simulation? *Philosophical Quarterly* 53 (211):243–255.

Campbell, Tom; Owhadi, Houman; Sauvageau, and David Watkinson (2017). On Testing the Simulation Theory. *International Journal of Quantum Foundations* 3: 78-99.

Chalmers, David J. (2018a). The Meta-Problem of Consciousness. Journal of Consciousness Studies, 25, No. 9–10, 2018, pp. 6–61

----- (2018b). Structuralism as a Response to Skepticism. *Journal of Philosophy* 115 (12):625-660.

----- (2017). The Virtual and the Real. *Disputatio* 9 (46):309-352.

----- (2005). The matrix as metaphysics. In Christopher Grau (ed.), *Philosophers Explore the Matrix*. Oxford University Press. pp. 132.

----- (1996). *The conscious mind: In search of a fundamental theory*. Oxford university press.

Dietrich, Eric (2011). There Is No Progress in Philosophy. *Essays in Philosophy* 12 (2):9.

Grange, Eric (2016). Mesh World P2P Hypothesis, https://www.delphitools.info/DWSH/, accessed Jan 10, 2019.

Huemer, Michael (2016). Serious theories and skeptical theories: Why you are probably not a brain in a vat. *Philosophical Studies* 173 (4):1031-1052.

Johnson, David Kyle (2011). Natural Evil and the Simulation Hypothesis. *Philo* 14 (2):161-175.

Leifer, M. S., & Pusey, M. F. (2017). Is a time symmetric interpretation of quantum theory possible without retrocausality?. *Proc. R. Soc. A*, *473*(2202), 20160607.

Le Poidevin, Robin (2015). The Experience and Perception of Time. *The Stanford Encyclopedia of Philosophy* (Summer 2015 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/sum2015/entries/time-experience/>.

Markosian, Ned (2016). Time. *The Stanford Encyclopedia of Philosophy* (Fall 2016 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/fall2016/entries/time/>.

Mizrahi, Moti (2017). The Fine-Tuning Argument and the Simulation Hypothesis. *Think* 16 (47):93-102.

Moravec, Hans (1998). Simulation, Consciousness, and Existence, https://www.frc.ri.cmu.edu/~hpm/project.archive/general.articles/1998/SimConEx.98.html, accessed Jan 11, 2019.

Ninan, D. (2009). Persistence and the First-Person Perspective. *Philosophical Review* 118 (4):425-464.

O'Connor, Timothy and Franklin, Christopher (2018). Free Will. *The Stanford Encyclopedia of Philosophy* (Fall 2018 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/fall2018/entries/freewill/>.

Padmanabhan, T. (1985). Physical significance of Planck length. *Annals of Physics*, *165*(1), 38-58.

Schaffer, Jonathan (2016). The Metaphysics of Causation. *The Stanford Encyclopedia of Philosophy* (Fall 2016 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/fall2016/entries/causation-metaphysics/>.

Slezak, Peter P. (2018). Is There Progress in Philosophy? The Case for Taking History Seriously. *Philosophy* 93 (4):529-555.

Whitworth, Brian (unpublished manuscript). 'The Physical World as a Virtual Reality', http://brianwhitworth.com/BW-VRT1.pdf, accessed Jan 10, 2019.

Willard, M. B. (2013). Game called on account of fog: metametaphysics and epistemic dismissivism. *Philosophical Studies* 164 (1):1-14.