 doi:10.20944/preprints202003.0072.v1

Article

Modelling Consciousness within Mental Monism: An Automata-Theoretic Approach

Peter B. Lloyd [1,*]

[1] School of Computing, University of Kent; peter@peterblloyd.com
* Correspondence: peter@peterblloyd.com; Tel.: +44-7856 675692

Abstract: Models of consciousness are usually developed within physical monist or dualistic frameworks, in which the structure and dynamics of the mind derive from the workings of the physical world (in particular, the brain). Little attention has been given to modeling within a mental monist framework, deriving the structure and dynamics of the mental world from primitive mental constituents only. Mental monism is gaining attention as a candidate solution to Chalmers' Hard Problem, and it is therefore timely to examine possible formal models of consciousness within it. Here, we propose a minimal set of hypotheses that any credible model of consciousness (within mental monism) should respect. From those hypotheses, it is feasible to construct many formal models that permit universal computation in the mental world, through cellular automata. We need further hypotheses to define transition rules for particular models, and we propose a transition rule with the unusual property of deep copying in the time dimension. In conclusion, we hope to dispel the notion that mental monism requires a *deus ex machina*, by showing that a parsimonious set of assumptions can yield a naturalistic and computationally potent mental world.

Keywords: idealism; consciousness; Hard Problem; automata theory; mental models.

1. Introduction

The modern mind-body problem, as formulated by Descartes [1], has expanded into the broad field of consciousness studies centred around what Chalmers [2] termed the 'Hard Problem of consciousness': even if we had a complete understanding of all physical processes in the brain, there would remain an explanatory gap between the physical workings of the brain and the operations of the conscious mind. Proposed solutions to this problem fall into three broad camps: physical monism, which holds that only physical things are real; dualism, which holds that both the physical and mental worlds are real, and that minds cannot be reduced to purely physical systems; and mental monism, which holds that, ultimately, reality consists only of conscious minds. Mental monism is very much a minority position, but has been receiving growing attention owing to the failure of the more orthodox schools of thought. On the one hand, physical monism, in effect, denies the existence of the very thing we are endeavouring to explain, and whose actual existence is witnessed in every moment of our waking lives, namely consciousness. On the other hand, dualism has to suppose the presence of a physical substrate that is inherently incapable of direct observation and whose existence and non-existence are operationally indistinguishable. Mental monism is the only class of solution that does not suffer those metaphysical failings. One of the things that has discouraged interest in this theory is the so-called 'bootstrap problem'. If there is nothing in reality except conscious minds, then the entire structure and dynamics of the mental world has to be explained by mental primitives, as there is no brain circuitry from which mental behavior could be derived; and, moreover, all facts and laws of the physical construct must likewise be derived from mental primitives. I call this the 'bootstrap problem' by analogy with computer engineering: as a computer must 'boot up' or metaphorically pull itself up by its bootstraps, when it is switched on, so the entire observed universe must boot up from elementary mental entities.

 doi:10.20944/preprints202003.0072.v1

Lloyd [3,4] has given a detailed defence of mental monism, and those arguments will not be rehearsed here. The starting point in this paper is to suppose that mental monism is true, and to consider what formal models of consciousness might work within that framework.

In the field of formal modelling of consciousness, attention is normally given to models that are situated within physical monist or dualist philosophical frameworks, and the structure and dynamics of the conscious mind can be based, either explicitly or implicitly, on the workings of the physical brain. Formal models within mental monism have been neglected. On the contrary, one frequently encounters 'straw man' objections to mental monism, asserting that it requires a *deus ex machina*, an intelligent deity that instigates and maintains the manifest world. Needless to say, such deistic accounts are explanatorily delinquent. The central take-home message of this paper is that mental monism does not require the *deus ex machina*, as it is possible to envisage computationally rich mental worlds bootstrapped from minimal assumptions about mental primitives.

2. Hypotheses for Possible Models in Mental Monism

Our method is to propose a number of plausible minimal hypotheses about the behaviour of the mental primitives, and show that this allows the creation of mental systems that are capable of universal computation. This is done within the framework of mental monism by make a formalized statement of mentalism as a prior result established by the present author [4]. It is argued that mental systems of this type are capable of embodying cellular automata, which are known to be capable of implementing Universal Turing Machines, which are formal automata systems that can compute any computable function.

These hypotheses under-determine the basic rules of the mental world. In particular, a key question concerns the transition rule obeyed by mental primitives when regarded as units in a cellular automaton. Traditionally, such rules are set by the investigator, on the premise that there is a substrate in which the cellular automata are implemented. It is argued that that would be an inelegant solution in the present case and unlikely to be fruitful. Instead, an unconventional kind of transition structure is considered to avoid a supposition of arbitrary transition rules, namely deep temporal copying.

Finally, we note the parallel between the cellular automata that are proposed on theoretical grounds by mental monism, and the cytoskeletal cellular automata, which Penrose and Hameroff have suggested as physical correlates of consciousness [5].

2.1. Hypothesis 1: Mental discreteness

A fundamental assumption in this approach is that the conscious mind is a discrete system, and therefore lends itself to the methods of automata theory. At the present nascent state of this field, it is by no means certain whether this assumption is correct, but the following remarks are intended to show that it is at least plausible, and to indicate that it is therefore legitimate to inquire into how this basic assumption might be developed.

The phenomenal field comprises discernible parts, which have spatial, temporal, structural, and qualitative relations, and are subject to limitations of acuity in space, time, logical composition, and quality. For example, I can perceive the contents of my visual field as a two-dimensional assemblage of discernible patches of different colours and brightness, and other features, with limits of resolution in mental time and space, limits of association with other mental content, and limits of discernible gradations of quality. In operational terms, the evidence for this can be outlined as follows. First, between any two points in any sensory field (visual, tactile, auditory, or proprioceptive), there is a finite number of discernible positions. Second, between the start and end of any interval of time, a finite number of moments can be discerned. Third, from any given experientia, there is a finite number of associated memories. Fourth, along any gradient of quality (brightness, redness, sweetness), there is a finite number of levels. We are concerned here, not with what the specific numbers are, but only with the fact that these intervals (spatial, temporal, associative, and qualitative) are not infinitely divisible, and that therefore a discrete system is a legitimate model to try out.

 doi:10.20944/preprints202003.0072.v1

There are some counter-arguments to this. It could be suggested that the conscious mind actually inhabits a continuous mental space, but that the contents happen to be discrete because they are produced in lock-step with discrete neural systems. A related argument is that the sensorium is actually continuous but we lack the skill to discern this continuity as we become vague and confused when mental content is too tightly packed. This could be so, but so far there is no evidence for it, and we are therefore entitled to begin by studying discrete models. Another counter-argument is that the apparent decomposition of the sensorium into finite numbers of discrete elements is not real but a methodological artefact, it is produced by the very method of trying to see what reality is made up of. According to this postmodern perspective, experienced reality has no genuine constituents, but we can create the impression of composition by performing certain operation on our mental content, so as to yield a decomposition. For example, on this view, the experience of a sticky toffee pudding is an unanalysable sense datum, and the process of decomposing that experience into specific flavours, smells, textures, heat feelings, sensation of solidity, softness, and stickiness, does not reveal the pre-existing constituents of the sense datum but only generates new sensations in isolation. Moreover the postmodernist would maintain that there are multiple, equally legitimate modes of decomposition, hence the decomposition cannot reveal a prior, objective composition, but merely produces a subjective derivative. This speaks to a deep methodological divide between the rational-scientific enterprise and the anti-rational and anti-scientific programme of postmodernism. The main effective argument against this line of thinking is utility: if we find that independent decompositions converge on a common model, and this model yields testable hypotheses that are found to be confirmed, then we can at least conclude that the model, *qua* model, is valid. The only way to make such an assessment is to attempt to model consciousness in this analytic manner.

As to whether these discrete products of the decomposition of mental wholes are real prior constituents, or only posterior artefacts, this might be confirmed only with enhanced skill in introspection. Meanwhile, we can at least say that mental discreteness is the more plausible hypothesis, and it is therefore legitimate as modelling hypothesis. Hence we are led to:

Hypothesis 1 (Discreteness): The phenomenal content of a conscious mind is a discrete system.

Therefore, a basic premise for modelling consciousness is that the phenomenal field can be represented as a time-varying structure $M(t) = \langle S, C(t), R(t) \rangle$ where S is the fixed subject, C(t) is a finite set of conscious experiences, R(t) is a finite set of binary relations between experiences, which we will define later; and t is discrete time.

In a classical cellular automaton, we have a set of unit automata, each of which has fixed spatial relations with its neighbours, and each possesses a number of states. In that conceptual framework, the unit automaton and its states are two ontologically distinct things; but, as we will see below, we will need a more parsimonious model than that for mental monism.

2.2. Premise 1: Mental monism

The hypothesis of discreteness warrants the following expression of the fundamental premise of this study, namely the philosophical theory of mental monism. Mental monism (also known as 'subjective idealism' or 'mentalism') is the doctrine that reality is wholly mental, and that what we take to be the physical world is a derived construct. This doctrine was proposed by George Berkeley in the Eighteenth Century [6], and articulated in modern terms by Foster [7], Lloyd [7,8], and Pearce [10]. Lloyd [4] provides a proof of this doctrine, roughly equivalent to those of Foster and Pearce, and following the original reasoning of Berkeley. It is acknowledged that mental monism is highly contentious, but the argument for it will not be rehearsed here as it has been adequately covered in the foregoing references. The task addressed in this paper is somewhat narrower: taking mental monism as a premise, how might we formally model the conscious mind?

Premise 1 (Mental Monism): reality consists just of a set of minds $\mathbf{U} = \{M_0, M_1, M_2, \ldots\}$.

It immediately follows from mental monism that physical space has no mind-independent reality and cannot serve as a ground within which minds exist. So, not only do we have to model the conscious mind without the substrate of the brain, we have to do so without a spatial medium in which to hold the conscious experiences. Hence:

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 March 2020 doi:10.20944/preprints202003.0072.v1

Corollary 1 (Non-Spatiality): Minds do not have physico-spatial relationships.

Furthermore, the conventional individuation of minds and subjects by means of their spatial separation is no longer feasible in mental monism as minds are not embedded in space. Conventionally, it is held that two minds are distinct by virtue of their sitting inside distinct brains, which are in different places. Likewise the subjects of those two minds inherit their individuation from the individuation of their respective minds. This has to be re-thought radically in mental monism. Thus two minds (as distinct from subjects) can be individuated by their content, but not by their position in space (which they don't have); whereas the subject has neither content nor position. Therefore, without containing content, and without being anchored in space, subjects simply have no individuation. There is therefore only one subject. (This argument is given in more detail by Lloyd [4].) Although Western thinking may view this conclusion as outrageous, the Hindu school of Advaita Vedanta has held it as a central tenet since the Eight Century CE: "Thou art Brahman". Hence:

Corollary 2 (Unified subject): There is at most one subject.

2.3. Hypothesis 2: Naturalism

We seek to account for the structure and dynamics of the mental world with a model that is nomologically constrained. For otherwise, without laws that govern the mental world, we have either chaos or magic, and not a usable model at all. As the mind is a dynamic system, the most natural form of nomological constraint is a causal relation, in which the successive states of a mind are largely driven by earlier states. This need not be a total determinism: a causal connection in which some state changes are nondeterministic could still be nomological enough to yield the stable behavour that we actually observe in the mind. Furthermore, we seek a model in which cause-and-effect works locally. The first thing we notice about the mental realm is that it is compartmentalized into minds, each of which is largely private and interacts with other minds through specific input/output channels. I am going to refer to this as 'naturalistic' because any alternative—that is, behaviour that is predominantly uncaused, or caused by events in other minds—would undermine our basic intuitions about how our minds work. At this point, we cannot prove this, but is a plausible starting point for modelling. So:

Hypothesis 2 (Naturalism): The time evolution of a mind is local, $M(t+1) \in \Psi(M(t))$, where the right-hand side is a set of possible mental states (possibly a singleton).

That is to say, given a state of a mind M at time t, a function Ψ applied to M determines the set of possible states at the next moment. In the real world, of course, a mind is affected by sensory inputs, but those inputs can affect the operation of the mind only after they have arrived 'inside' the mind. We will discuss below the problem of how inputs get into the mind, but for now let us just note that the standard physical-realist route is not available in mental monism. Specifically, the physical-realist route is as follows. In any model where the mind supervenes on the brain, a physical stimulus physically arrives on the doorstep of the brain (for example, light waves arrive at the retina, and the normal local physical processes—that is, signal propagation along nerve fibres—carry that stimulus into the brain, where the supervenience is supposed to happen). In such a framework, mental input piggybacks on the physical proximity of sensory stimuli to the neural correlates of consciousness. In mental monism, there is no space, hence no proximity, hence input into the substrate-free mind must be modelled quite differently from the physical-realist route.

2.4. Hypothesis 3: Volition

For some people, an hypothesis of volition is somewhat extravagant and should not be assigned to a minimal set of hypotheses. I suggest, however, that volition is, in fact, a feature of the conscious mind that is universally apprehended and must be modelled as a fundamental component as a nondeterminism cannot be modelled by a deterministic system.

Hypothesis 3 (Volition): The time evolution of the mind involves a nondeterministic component. That is, the time evolution, Ψ, may yield a non-singleton set of outcomes, $|\Psi(M(t))| \geq 1$. Free will determines which outcome is selected: $M(t+1) = \omega(\Psi(M(t)))$.

2.5. Hypothesis 3: Unitary experientiae

We now turn to the transmission of causal effect, and hence of information, between minds, and between parts of a mind. In physics, we are accustomed to the central role that proximity plays in the transmission of causation. The classic example is a billiard ball striking its neighbour and imparting kinetic energy; a photon striking the retina and imparting its electromagnetic energy; a gravitational field gripping an object and pulling it in. Even non-local quantum-mechanical phenomena do not involve the observable transmission of energy or information from A to B instantly without passing through the intervening space. In mental monism, this role of proximity cannot be utlilised, as there is no concept of proximity, because (by Corollary 1) there is no space. So, we have to rethink the mode of causal transmission in mental monism.

At a first inspection, it seems that the mind comprises both experiences ('ideas' in the 17th Century term used by Locke and Berkeley) and acts of volition—the latter manifested both as changes in internal thoughts and imagery, and as motor actions. That apparently basic conceptual separation between experiences and volitions is, I shall now argue, not possible in mental monism. This counter-intuitive conclusion is, I shall argue, an unexpected consequence of the non-spatial mode of causal transmission in the realm of consciousness.

We shall consider in turn 'outward perceptions' and 'inward imaginations'. (As discussed by Lloyd [4] and Pearce [10], 'outward' is used in its everyday sense without indicating anything literally outside the mind.) Let us consider the following supposition, and then show that it leads to absurdity, and must therefore be rejected.

Supposition 1: A conscious mind comprises elements of two distinct classes: experiences and volitions.

In the first line of argument ('Outward perception') only, we shall also consider, and reject, the following supposition:

Supposition 2: A volition in one mind produces experiences in another mind through an intermediate mechanism.

Outward perceptions. Consider, first, perceptions that occur in your mind due to the manifest world. For example, you look up at the sky and have an experience of phenomenal blue. By Hypothesis 2 (Naturalism), this is not random noise but the product of some process. By Premise 1 (Mental Monism), that process must be the activity of a conscious mind, since there are no other entities that could be producing the phenomenal blue. Following Lloyd [8], we will use the general label of 'metamind' for the mind(s) responsible for natural phenomena, that is, all actions that are not acts of personal volition. This term is preferred over Berkeley's "God" or Shankara's "Brahman" as it avoids religious connotations. Therefore, we may say that your sensation of phenomenal blue in the sky is produced by a volition of the metamind. At first, it may seem that the metamental volition and your phenomenal experience are two distinct things, and we must therefore suppose there is some intermediate mechanism that transmits the causal influence from the volition to the perception, for Hypothesis 2 (Naturalism) prohibits any magical remote influence. What could be the nature of that mechanism? By Premise 1 (Mental Monism) reality comprises only conscious minds. So, the putative mechanism that conveys causality from mind A to mind B, must be a third mind C. And, by our target Supposition 1, mind C consists of experiences and volitions: which leads us to a circularity, because we must explain the working of mind C by supposing that the volition of A produces an experience in C, which by the foregoing logic entails an intermediate mechanism from A's volition to C's experience, which must be another mind, D. It is clear that the supposition of an intermediate mechanism and the composition of minds by distinct experiences and volitions leads us to absurdity.

We might be tempted to think there could be an articulation between the agent and percipient (in our example, the agent is the metamind and the percipient is your personal mind). For example, could it be that the agent is 'adjacent' to the percipient in some sense, and just rubbed up against it in some causally efficacious way? No, because, by Corollary 1, we have no concept of spatial distance between minds, hence no concept of adjacency of minds. So the agent cannot affect the percipient through 'contact'.

 doi:10.20944/preprints202003.0072.v1

By this process of elimination, the only concept left on the table is that the agent overlaps with the percipient. So, the metamental volition occurs *inside* the percipient mind. Thus Supposition 2 must be dismissed, and we can now focus on Supposition 1.

Inward imagination. This also brings us to the second case, that of an intra-mental act of volition, for example, where you imagine something. We are familiar with our own imaginings, but by the foregoing argument, even outward perception must be understood as an intra-mental act of volition (as the metamind's volition to produce the blue patch must occur inside your mind). The following remarks therefore apply to both cases, outward and inward.

Now we have an analogous logical conundrum to the one that was addressed above. If we suppose that the causal influence between the volition and the experience (within a mind) is conveyed by an intermediate entity then the intermediary can be only another volition or an experience, which just leads us to the absurdity of an infinite regress. And if we were to suppose that the proximity of the volition to the experience would enable a causal power to be transmitted, then we are stuck again as there is no intra-mental space, any more that there is any inter-mental space. Now this process of elimination leaves us only one option on the table, that the volition is inside the experience. But these are elementary things, so we are forced to conclude that the volition and the experience are the same thing. I shall use the term 'experientia' for this unitary entity that is both volition and experience.

Furthermore we must say that every experience is also a volition (for, otherwise, what is there to cause it?) and every volition is an experience (for, otherwise it would not produce any result).

This brings us to a simple picture in which reality comprises a countable set of minds, and each mind comprises a countable set of experientiae, each of which is both volition and experience, and one mind communicates with another by executing a volition within the recipient mind. The agent mind and recipient mind each may be personal minds or the metamind.

Result 1. The contents of a conscious mind is a set of unitary experientiae. Each experientia is both volition and experience and, conversely, every volition and every experience is an experientia.

Result 2. Communication between two minds is achieved by executing a volition inside the recipient mind.

I admit that these two results seem implausible, but I submit that this is just because we are so accustomed the spatially grounded causality of physics, and we have to shift to a way of thinking that is closer to that used in computer science.

2.6. Hypothesis 4: The construction of mental space

Our experientia are arranged in a space-like configuration. Within physicalist or dualist models, this can readily be explained by referring to to the underlying brain structure, which is spatially isomorphic to afferent nerve endings. Within the framework of mental monism, however, we do not have that substrate, and mental space must be constructed out of mentations only.

What are the phenomenal elements, and what operators act on them? On a naïve view, we may be tempted to conceive of the sensorium as being like an inner cinema screen. A moment's reflection shows that this is untenable as it implies a homuncula watching the inner cinema screen, who in turn has her own inner cinema screen, *ad infinitum*. On a first inspection, it would appear that there is a persistent body-image, consisting of a finite number of 'places', each of which can contain sensory experiences of a type specific to that section of the sensorium. For example, my skin appears to form a two-dimensional space, topologically equivalent to a sphere, divided up into places each of which can harbour sensations of heat, texture, pressure, and so on.

It is therefore tempting to envisage the manifold of mental places as being like an array of pixels, but that does not sit well with experimental data. Since the work of Hubel & Weisel, it is apparent that the ingredients of the visual sensorium include lines and movements as well as points. Nevertheless, we can use the term 'places' for these spatially differentiated components, as long as we bear in mind that they are not just point-like 'mental pixels'.

With regard to outward-facing senses, this architecture of mental space is self-evident: your visual experiences occur in a two-dimensional field of vision. You automatically project what you see

 doi:10.20944/preprints202003.0072.v1

into the constructed three-space surrounding your body, but your visual field is not itself three-dimensional, it only has associations of three-dimensional structure attached to it. Extending sideways from the visual field is the tactual field. Although this is a different sensory modality from sight, it is situated in the same mental space, a fact that is trivially established by touching your eye: you see the finger approach the pupil, and then you feel the eyeball embedded within the tactual field. Likewise hearing: although the source of a sound is more diffusely spread around the circumpersonal space, the auditory experientiae are situated in the same mental space as vision and touch. This is trivially established when you put your fingers in your ears to block out a sound. We can go through all the senses in like manner: taste and smell are situated in the body-image at the buccal and nasal passages. Proprioception locates qualia of limb position in intermediate positions within the topological boundary of the body image. Pain sensations such as nausea, toothache, headache are likewise situated inside the body image. Thoughts are experienced as if contained within the body-image of the head and, if articulated will be 'heard' as vocal sounds in the 'inner ear' (that is, projected to the positions of the ears) or 'seen' as words in the 'inner eye' (that is, projected to the visual field). Braille users project words to tactile sensations in the fingers. Emotions are felt in the head or in organs of the body (for example, fear in the intestines). Although thoughts and emotions carry a propositional freight, our immediate awareness of them is constituted by qualia. You know that you have a certain thought only by virtue of the auditory, visual, or other sensory markers that indicate the thought. (Maybe, when reading this paragraph, your heard the words "That's rubbish!" in your inner ear? Or saw yourself writing "Thoughts have no qualia"?) No experience is wholly without a spatial relation to the rest of the mental contents. All mental content is situated in the mental space of the body image.

What, precisely, are those spatial relations? Recall that, in mental monism, there is no given physical space within which experientae could relate spatially. Mental spatial relations must therefore be wholly constituted by some pre-spatial aspect of the mental structure and dynamics. So, we must inquire what precisely it means to say, for example, that a particular green patch in the visual field is below a particular blue patch. The naïve answer would be that the two patches are situated at two positions in a spatial medium, and the vertical coordinate of one position is greater than that of the other. That naïve account, however, will not work in mental monism because there is no prior space in which experientiae are sitting. Suppose there were such a space. Then it would have to be an experientia. And then we would have to explain what it means for the blue-patch experientia to be in a certain position in relation to the space experientia. Which would require a further space in which the space experientia sits, which begins an infinite regress. For sure, whatever it is that constitutes spatial relations must comprise experientiae, as those are the only building blocks of the mind allowed by mental monism. But they could not be static, for example, an 'aboveness' quale and a 'belowness' quale, as that would underdetermine the direction of the relation: if I had a blue patch, a green patch, and a belowness quale, then would that constitute green's being below blue, or vice versa? Rather, spatial relations must be determined by a dynamic process. If we operationalise the mathematical concept of space, we see that it is the capacity for movement. That, I suggest, points to the correct model of spatial relations in the sensorium.

My hypothesis is that the remembered pattern of movements and consequent changes of perceptual content constitutes the perception of space. Certainly, the commutative relation of movement provides a sufficient basis for the commutative relation of position. Thus if T, G, and R denote three qualitatively different sensations, and u and d denote two opposite movements (say, upward and downward) then the sequences <T,u,G,u,R> and <T,d,G,d,R> are sufficient to define the spatial relation of those qualities in the sensorium. When sensations are assembled into a topology in this kind, we refer to the resulting construction as the body image. I suggest that the memory of these volitional relations of this kind is constitutive of the apparent spatial relations: two sensations have a spatial relation by virtue of their being embedded in a body image, which is constituted by a network of remembered volitions and changes of experience.

I shall outline two examples. The first is chosen for its simplicity, which lets us introspect more clearly the principles at work; the second is chosen because it has some empirical support. First,

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 March 2020 doi:10.20944/preprints202003.0072.v1

consider the buccal cavity, that is, the interior of the mouth. Unlike of the rest of surface of the body, we do not habitually inspect the mouth either visually or digitally: although, of course, we can and occasionally do look inside the mouth with a mirror, or insert our fingers, predominantly we do not. Instead the "conscious mouth image" is formed from the somatosensory experiences of the only movable piece of anatomy inside it, namely the tongue [11]. Press the tip of your tongue against the inside of your incisor teeth (sensation T), now move it upward and gain a different sensation of the gum (G), and move it upward again to the roof of the mouth (R); then reverse the movement. The memory of these two sequences, <T,u,G,u,R> and <R,d,G,d,T>, form part of the body image of the mouth. The whole mouth image, and hence all spatial relations of mouth sensations, are, I submit, constituted in this manner.

Second, suppose I am looking out from my garden table (T), and I see a green field (G) and the red sunset (R) above it. (It matters not whether this is a waking experience, or a dream, or an hallucination). If I move my gaze upwards, then I find that the foveal part of the visual field, which was brown becomes green, and after a further movement becomes red; and if I move my gaze downwards then my fovea shows green and brown again. The memory of these two sequences, <T,u,G,u,R> and <R,d,G,d,T>, and a myriad similar ones, constitute the spatial skeleton of visual field. Saccadic movements of the eye provide a constant flux of minute movements of the eye to and fro, maintaining the visual space. On this view, if the saccades were to cease, the visual space would disintegrate. You can approximate this condition for peripheral vision simply by staring fixedly into your own eyes in the mirror for several minutes, and observing the perception disintegrate and be replaced by imaginings [12], a process related to the Troxler effect.

The foregoing does not prove the dynamic model of mindspace, but it does justify it as a plausible hypothesis, hence:

Hypothesis 4 (Mindspace): Spatial relations of experientiae in the conscious mind are constituted by remembered sequences of volitions and consequent changes of experiences.

For two experiences E_A, E_B, we will say they are proximal *prox*(E_A, E_B), if there is some sequence of volitions V_1, V_2, …, V_n and experiences F_1, F_2, … F_{n-1} such that E_A, V_1, F_1, V_2, F_2, … F_{n-1}, V_n, E_B is a remembered compound action, and if there also exists a sequence of volitions W_1, W_2, …, W_n such that E_B, W_1, F_{n-1}, W_2, G_{n-2}, … F_1, W_n, E_A, is a remembered compound action. Obviously this is commutative, *prox*(E_A, E_B) iff *prox*(E_B, E_A). Furthermore, two experiences E_A, E_B will be spatially related, *spat*(E_A,E_B), if either *prox*(E_A,E_B) or there is some sequence of experiences G_1, G_2, G_p, such that *prox*(G_i,G_{i+1}) for i = 1 to p-1. Obviously this is commutative, *spat*(E_A, E_B) iff *spat*(E_B, E_A).

Let me emphasise that this is not supposed to be a proof that this is what mental space really is. The point, rather, is that if mental monism is true, then mental space must be built up from mental primitives in something like the above manner. The foregoing is proposed as a plausible possible account of mental space: I expect further research to discover more accurate models of *prox* and *spat*. Going forwards, we may assumed that *prox* and *spat* exist and are built up from experiences and volitions in some manneri or other.

2.7. Hypothesis 5: Mental individuation and intermental portals

Recall the distinction between a conscious mind and its subject, the latter being the featureless agent of perception and volition in the mind. We saw above that, without spatial location, subjects cannot be individuated, and therefore what appear to be distinct personal subjects are, in fact, numerically identical. Minds, on the other hand, can be individuated by their content, and we will now examine how this could be implemented.

It is a matter of everyday experience that two minds M_1 and M_2 are mutually private but can nonetheless communicate. In physical-realism, this is straightforward to characterise and explain. Each mind M_1 is spatially enclosed in a brain, which has afferent and efferent nerve fibres that allow communication with a shared environment, which serves as a medium for this mind to transact an exchange of information with M_2 and other minds. In mental monism, it is not so easy: there is no space in which to contain and isolate minds, and there is no non-mental medium through which minds can communicate. Instead we must conceptualise privacy and intercommunication within the

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 March 2020 doi:10.20944/preprints202003.0072.v1

sparse ontology that is dictated by the theory of mental monism. Therefore, we will first consider a model of mental privacy, using the ideas developed above, and then adapt that to model intermental communication.

(a) Individuation. In mental monism, there is a universal set of experientiae that is the union of all personal minds' sensoria, $U_c = C_1 \cup C_2 \cup \ldots$ What partitions this universal set into personal subsets? In other words, what constitutes the boundaries of a personal mind? An inspection of your own sensorium right now will reveal, I believe, that all of your experientiae are situated within a personal mental space—which, whilst it is not in physical space (although it forms the seed for your intuition of three-dimensional space in the physical construct). In everyday experience, you find that you can move your attention to any area of your sensorium, thoughts, and imaginings, and can take note of that content, and act upon it, and you can initiate volitional acts to change your thoughts and imaginings and make movements of your body. You can associatively retrieve memories into attention, and lay down new memories, but your sensorium, thoughts, and memories are accessible only by you. In operational terms, therefore, it appears that your mind is closed under operations of access, and this is constitutive of mental individuation and the boundaries between minds.

This might seem odd for thoughts and emotions but, as we saw above, under the austere ontology of mental monism, all mental contents exist only as qualia, and hence tied to a sensory modality, and hence embedded in mental space.

As we saw above in Hypothesis 4, the mental space is woven from volitions linked to experiences, and we may therefore suppose that the boundary of the personal mind is constituted by their limit. Thus, two experientiae E_A and E_B are termed co-mental, *co-m*(E_A, E_B), if they belong to the same mind. Building on the concept of mental space, we can now hypothesis that the edge of mental space marks the partition of U into C_i. Thus *co-m*(E_A, E_B) iff *spat*(E_A, E_B).

(b) Communication. If two minds M_1 and M_2 are to interact then, as we have noted above, they must carry out their inter-mental communication in an intersection of the contents of two minds, which I will refer to as a 'portal'. $P_{i,j} = C_i \cap C_j$ where $M_i = \langle S, C_i, R_i \rangle$. If a mind M_1 changes part of its contents that is not within the portal, then another mind M_2 will not know about it. Only if M_1 changes something within the subset that constitutes the portal can M_2 detect it.

Within the portal, a change of content that is executed by one mind is deemed to be its 'output', and when it is detected by the other mind, it is that one's 'input'. In principle, a portal is bidirectional: the two minds who share it could use it for input and output. In practice, portals are unidirectional, at least in complex organisms such as ourselves. The reason for this is clear. The output from a mind—which corresponds to motor activity—is not random noise but comprises intentional acts that are 'pre-processed', that is planned and controlled by computational machinery sitting outside the portal. Likewise the input into a mind is 'post-processed', that is analysed and cognised, by other computational machinery. There are no 'walls' between one portal and another, nor between a portal and the rest of the mind, but the notion of 'a portal' makes sense if its pre- or post-processing is a functionally distinct part of the mind. (A percipient mind would not normally change the content of its own input portal, but doing so would be classed as dreaming or hallucinating.)

Tying together these ideas, we have:

Hypothesis 5 (Interface): The contents of a mind are closed under operations of access, the formal relaton spat forming a mental space. Communication between minds occurs through intersections between the communicating minds.

A mind could have multiple portals with another mind, and there is no reason they could not overlap. Although the hypothesised portals model any direct communication between minds, in fact almost all of the direct communication that a personal mind engages in is with the part of the metamind that drives that mind's avatar in the physical construct. Our sensorimotor engagement with our surroundings is mediated by the elements of the metamind that manifest as sense organs and muscles of the avatar. Likewise, the pre-conscious operations of the mind, the storage and retrieval of memories, the recognition of faces, the cognition and construction of sentences, all rely upon interaction with the metamental units that govern the brain and nervous system. The metamental portals in the personal mind are so extensive that the personal mind has little more than

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 March 2020 doi:10.20944/preprints202003.0072.v1

a central executive role to call its own. Be that as it may, our concern here is to model the basic mechanisms involved.

In the models of Hoffman [18] and Kastrup [21], the boundary of the mind is considered as a Markov blanket. The model proposed here would be consistent with that conceptualisation.

2.8. Hypothesis 6: Elemenatary experientiae

The phenomenal contents of the mind exhibit a rich qualitative variety (colours, flavours, emotions), and it would be surprising if these were all primitive elements of reality. Our reductionist instinct makes us seek to explain them as composites of some more basic constituents. Mental monism, however, excludes the existence of non-phenomenal elements: if something cannot be experienced by a conscious mind then it cannot really exist. That does not, however, mean that the basic constituents must always be noticed by the percipient. (This distinction reflects Block's differentiation of phenomenal and access consciousness [13].)

We can find examples of this phenomenal composition in subtle sensory experiences such as the sound of orchestral music, the taste of whisky, or the colours of a painting. A trained mind can analyse a compound perception into its constituent phenomenal elements, which the untrained mind cannot notice. The unskilled palate can differentiate Laphraoig from Macallan without being able to articulate any of the component flavours of that difference. This does not mean the constituent sensations are not present in phenomena consciousness: they are there, but the person does not have the wherewithal to pick them out and recognize them.

Therefore, it is at least a coherent proposal that the rich repertoire of our experientia is constituted by some small set of phenomenal primitives that are capable of being apprehended consciously but are normally beyond the liminal boundary. That primitive set could even be a singleton: it is at least a coherent hypothesis, albeit surprising, that there is a single elementary experientia out of which all phenomenal contents of the mind are formed.

Hypothesis 6 (Elementary Experientia): All experientia resolve into a small set of 'elementary experientiae', which are phenomenal primitives that are capable of being consciously experienced, and that admit of no further analysis.

2.9. Hypothesis 7: Elemenatary operators

I have proposed the 'elementary experientia' as the basic constituent of the mind. What are the minimal operators we need to posit for these elements?

In the ontogenesis of a mind, the new mind starts from very little, or nothing at all, and ends up with vast numbers of sensorium places. So, as a minimum, we need a creation operator; and, for symmetry, we might suppose a destruction operator. If we suppose the simplest form of experientia is a bare phenomenal existent, then the minimal operators are a correspondingly simple pair that merely create—in a mitosis-like fission, leading to two identical phenomenal cells—and merely destroy in an annihilation. Whatever the actual elementary experientiae and operators turn out to be, it is hard to imagine they would not include these bare minima.

Unlike the mitosis of an amoeba, the daughter cell cannot float free of the parent cell, as there is no spatial medium in which to float off. And if anything did, *per impossible*, float off, it would simply be lost. We may think of the daughter cell as inhering as a feature of the parent, rather than setting off as a new existent. Does the parent cell automatically get annihilated after creating the daughter? The simpler hypothesis is that nothing happens to the parent: it simply carries on until an annihilation operator gets applied to it.

We thus have a picture of a chain of experientiae comprising successive daughter cells $\langle m_1, m_2, m_3 \ldots, m_n \rangle$, which can grow further through the fission of the terminal cell m_n.

Again for the sake of simpler hypotheses, we will suppose that there is nothing prohibiting non-terminal experientiae from creating daughters. That is, we will suppose that an experientia can undergo fission even if it is not the final cell. In this way, the chain becomes branched like a tree. Let us suppose that a nonterminal cell m_i first yields a chain $\langle m_i, m_{i+1}, \ldots, m_{i+n} \rangle$. If an intermediate cell m_{i+k} undergoes fission, where $0 < k < n$, then we can consider two possible models, either a budding

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 March 2020 doi:10.20944/preprints202003.0072.v1

(<m_i,m_{i+1}, …., <m^1_{i+k},m^1_{i+k+1}, ..., m^1_{i+k+p}>, …., m_{j+n}> where p is initially 1) or a branch duplication (<m_i,m_{i+1}, …., m_{i+k-1}, <m^1_{i+k},m^1_{i+k+1}, …., m^1_{i+k+n}>, m_{i+k} , … , m_{i+n}>). In computing terms, the contrast is between a shallow copy or a deep copy.

The weak anthropic principle can help us in selecting a plausible model. For, we know that we live in a stable and complex world, so models that don't have the resources to create such a world can be excluded from consideration. Now, whether reality involves shallow or deep copying is an open question, but as I will argue below, a universe based on shallow copying of experientiae is unlikely to have yielded a stable world, unless we bring in some rather arbitrary transition rules. A further hypothesis that enables a stable world to develop is that, in a deep copy, siblings remain neighbours. Thus if our chain <m_i,m_{i+1}, …., m_{i+n}> fissions at k to yield (<m_i,m_{i+1}, …., <m^1_{i+k},m^1_{i+k+1}, …., m^1_{i+n}>, <m^2_{i+k},m^2_{i+k+1}, …., m^2_{i+n}>, m_{i+k} , … , m_{i+n}> then m_{i+k+j}, m^1_{i+k+j}, and m^2_{i+k+j}, are neighbours for j = 1 to n.

A third hypothesis that is motivated by the weak anthropic principle concerns the transition function. In line with Hypothesis 2 (Naturalism), we may assume that the fission operator behaves orthonormally, which might be partly or wholly stochastic; and we may suppose that the fission operation is driven by its neighbourhood of experientiae. That is, whether a given experientia annihilates, stays unchanged, or divides into a daughter, is governed by a transition rule that may be a mix of determinism and randomness, referring to its neighbors.

Hypothesis 7 (Operators): There exist two operators acting upon individual experientiae: creation and annihilation. The creation operator acts upon an experientia to make an identical copy of the experientia and all its descendants. Each descendent will become a neighbor of its counterpart in the created chain of new descendants. The annihilation operator permanently destroys an experientia. The application of these operators at time t is a function of the pattern of existing neighbours at t-1.

Which experientiae count as the neighbours of a given experientia is an open question. We have a wide range of possibilities. One such possibility is that only the siblings are effective as neighbours. If there are multiple deep copies made in rectilinear manner, so that m_i fissions to m^1_i, which repeats the exercise to become m^p_i to m^r_i, and if each m^q_i also fissions into an orthogonal chain $m^{q,1}_i$, $m^{q,2}_i$,… then we have series of infinitely extensible two-dimensional grid-like layers. Each 'cell' in the plane functions as a unit automaton, which can be occupied by an existing experientia or be empty, and can change between those two conditions in accordance with transition rules based on the neighbours. In this specific case, we have a classical two-dimensional cellular automaton [14].

It is a standard result that some binary-state cellular automata are capable of functioning as a universal Turing machine, and could therefore be capable of computing any computable function [15]. As this planar cellular computer could be produced by initiating deep copying from any starting point, we have a hierarchical computational structure that constitutes an object-oriented architecture. In fact, it has been shown that one-dimensional cellular automata can implement universal computation, albeit with a more than binary state space. This was conjectured by Wolfram in 1985, and a proof by Cook was presented in 1998 and published in 2004 [15].

Of course, the suppositions involved in producing this classical cellular automaton are deliberately contrived to yield that result. Quite different facts may obtain in nature. The point of the exercise, however, is to offer an existence proof: we have shown that this model of the basic units of conscious experience encompasses at least one instantiation of a universal computer. The model does, however, provide a vastly richer space of possible instantiations, some others of which might enable universal computation.

2.10. Hypothesis 8: Stochastic transition

The actual attributes of elementary experientiae, and of their transition functions, are completely unknown at this stage. I have argued above that the operators on elementaty experientiae will include creation and annihilation. We do not yet have tools to access the elementary mentations in order to study them scientifically. Our twin starting points will be: (a) computer simulations of possible

 doi:10.20944/preprints202003.0072.v1

permutations of the basic model; and (b) studying the logical structure of the neural correlates of consciousness (which we don't know for sure yet).

One of the components of the model to be settled is the transition function. In conventional studies of cellular automata, we define the transition function *ab initio*. For example, in John Conway's widely studied *Game of Life* [14], the rule is that a unit automaton is created when an empty cell is surrounded by three other unit automata, but annihilates if surrounded by zero or four. There is, however, a question of inelegance here. Why would the universe, at its elementary level, be governed by a rule as arbitrary as Conway's? From a subjective point of view, it would seem more elegant to avoid prescribing any particular deterministic transition. Instead, we want to suggest a wholly stochastic mode of operation, for example, that an experientia has a fixed probability of being annihilated (say 50%) at the next step, and if it persists then it has a fixed probability (say 50%) of breeding through deep copying of descendants). At some level of annihilation probability (not necessarily 50%), the statistical expectation is that the system would survive and grow, as opposed to fizzling out. Whether a purely stochastic transition rule would actually yield interesting behaviours, at least one capable of universal computation and self-reproduction, is unknown. We might have to consider such exotica as retroactive transitions, or transition rules that can change in time.

Hypothesis 8 (Stochastic transitions): The transitions of elementary experientiae are wholly stochastic.

We should emphasise that we have no philosophical or empirical grounds for this particular hypothesis: it is put forward only for reasons of internal elegance.

Unlike any physicalist or dualist model, the mentalist model allows conscious mechanisms to exist without any physical correlate, except for the boundary condition of compliance at the point of neural correlation.

3. Discussion

As mental monism is a minority position, there are correspondingly few writers who have addressed the modelling of the conscious mind from that position. I will briefly compare and contrast the present model with the proposals of Donald Hoffman and Bernardo Kastrup.

3.1. Donald Hoffman

Although he makes very little reference to Berkeley, Hoffman [16,17] has proposed what is essentially a mental monism theory, albeit one in which the physical construct arises from interactions among a community of equipotent 'conscious agents' as opposed to the more orthodox Berkeleian theory in which a central operating system (Berkeley's 'God') imposes the construct upon the community of personal minds. Here, I will not attempt to review Hoffman's extensive writings, only to compare Hoffman's basic theory with what is proposed in this paper.

The term MUI (mental user interface) was introduced by Lloyd [9] to denote a view of the physical construct that enables a conscious mind to have a single structure with sensory content and and volitional handles, through which it can interact with whatever 'external world' exists outside the personal mind. The analogy with a GUI (graphical user interface) is obvious. Hoffman [17] independently reintroduced the term MUI (multimodal user interface) for the same idea, and his theory covers both the mechanism of the MUI and the evolutionary development of it.

Hoffman begins his theory at a higher level than that considered here. His basic unit is the 'conscious agent', and he does not model the constituent components and their operations, which has been the focus of the present paper, namely the hypotheses about experientiae.

Hoffman [16] uses the same term for the conscious mind and for its mathematical model. On p 188, he writes that a conscious agent "perceives, decides and acts", which are actions ascribable to conscious minds. But on the same page he says that a conscious agent contains measurable sets, but a measurable set is a 2-tuple <S,M> where M is a subset of the power set of S, $S \subseteq \mathcal{P}(S)$, which means that a conscious agent is in the ontological class of mathematical abstractions. Even if the elements of S are ingredients of a real-world mind, the elements of M are artefacts of the description of S. Hoffman

confirms this further down the page: "The definition of a conscious agent is just math. … the mathematical model of conscious agents is not, and cannot create, consciousness." Here, let us disambiguate his nomenclature by using 'actual agent' for the actual conscious agent and 'model agent' for the mathematical model of the actual agent. This is not a semantic haggle but a substantive problem, as will become apparent. The model agent contains a measure M and the modeler is at liberty to define M any useful way. In contrast, the actual agent has no measure: there is no fact of the matter whether the actual mental elements have any particular set of subsets. (Otherwise, Hoffman would have to propose that sets are actual constituents of reality, alongside experiences and volitions. That, however, would be a move away from mental monism, which seems to be a fundamental tenet of Hoffman's work.)

Hoffman needs to introduce measurable spaces into his theory only because he allows the contents of the mind to be continuous. For the power set of a continuous space includes many 'perverse' non-measurable sets, but probability theory applies only to measurable spaces. In a finite, discrete space, the power set of the space is perfectly adequate as a basis for applying probability. In the present paper, the hypothesis of discreteness thus allows us the economy of not bringing in measurable spaces. At the end of the day, however it will be an empirical question whether the mind is discrete or continuous.

He then asserts that "every aspect of consciousness can be modeled by conscious agents", but he excepts qualia, regarding which he already asserted on p 44, "I will avoid this term because it often triggers debates about its precise definition. I will instead refer to conscious experiences." Like the present paper, only the structure and dynamics of the mind are modeled. Hypothesis 6 (elementary experientae) touches on this, but a full model of the qualitative content of qualia is a lacuna in both Hoffman's theory and the present paper.

Hoffman allows conscious agents to be minimal, but stops short of modelling their elements: "There is a bottom to the hierarchy of conscious agents. At the bottom reside the most elementary agents—"one-bit" agents—having just two experiences and two actions. The dynamics of a one-bit agent, and of interactions between two such agents, can be analyzed completely."

The full definition of Hoffman's model agent is: "A conscious agent, C, is a seven tuple C = (X, G, W, P, D, A, T), where X, G, and W are measurable spaces, P: W × X → X, D: X × G → G, and A: G × W → W are Markovian kernels, and T is a totally ordered set" [16, p.203]. The intention is that X represents the conscious experiences, G the actions, W the outside world. A 'Markovian kernel' is a stochastic transition function, so P represents perception of the world, D represents decisions, and G represents actions. T is discrete time. Hoffman's hypothesis is that the world, W, consists of conscious agents. Activity with and between model agents is represented by the Markovian kernels, but Hoffman does not model those functions: there is no explanation of why the kernel are such and thus, or what mechanism makes them tick.

Hoffman shows that a space can be constructed on the back of this formalism, and suggests that this could be a model of physical spacetime. Here, however, we return to the basic flaw that Hoffman conflates the map with territory: the modeler can defined whatever spaces she wishes on an underlying set, but the sensorium of a personal mind has an actual mental space (e.g. your visual field). The latter needs to be modeled in terms of elementary mental components, but Hoffman does not connect the two. He does not given an account (within the mental monist ontology) of how a blue patch can be above a green patch in your visual field.

What Hoffman has formulated is essentially a very general formalism, which needs to be filled out with a specific theory. What I have proposed is a specific theory, but one that is committed to a discrete space, and hence lends itself to the formalism of automata theory, but does not necessarily contradict Hoffman's formalism.

3.2. Bernardo Kastrup

Kastrup [19] explicitly advocates an idealist philosophical position, and acknowledges the need for idealism ultimately to ground the whole of physics. Nevertheless what he offers in this direction are metaphors rather than mathematical models. On the one hand, the community of personal minds

 doi:10.20944/preprints202003.0072.v1

is said to be like the alters of a cosmic Dissociative Identity Disorder [20]. On the other hand, the contents of a mind are supposed to arise from interference patterns as if in some wave-bearing medium: "experience is a pattern of excitation of TWE [that which experiences, i.e. the subject]" [20] But no such medium can exist because in mental monism the mental contents are all we have, so his 'patterns of excitation' cannot be more than a metaphor.

3.3. Evolution

Evolution is central to Hoffman's general approach. Kastrup also touches on it. Opponents of mental monism often attack a Berkeleian straw man, supposing that mental monism requires a magical *deus ex machina* that creates reality by intelligent design, and operates it by whimsical *fiat*. Such a theory would be explanatorily delinquent, as its starting point is a divine entity that is already at least as complex as the world we are seeking to explain. It is not a theory that was ever advanced by major idealists such as Shankara or Berkeley.

It does, however, highlight that mental monism owes us an explanation how the world as we know it comes to be. We suggest that such an explanation might be formulated within the automata-theoretic approached sketched out here.

Consider the primordial soup of phenomenal elements before the formation of the physical construct. A structure of experientiae that acquires a capability for persistence will persist while others than have a purely ephemeral formation will disappear. Persistence means the persistence of a volitional structure. Suppose that, within that structure, a sub-structure chances to arise that can not only persist but also reproduce itself. By Darwinian pressure, it will prevail over the rest of the primordial soup. If such as structure can build a physical construct with the fundamental elements of physics such as quantum mechanics, then the bootstrap is complete: the rest is history.

3.4. Cytoskeletal cellular automata

Penrose and Hameroff [5] have hypothesized that the physical correlate of consciousness in the brain is the microtubule. The philosophical theory of mental monism excludes the Penrose-Hameroff philosophy of identifying qualia with the objective collapse of quantum superpositions in the microtubule. Nevertheless, their arguments are attractive for regarding the microtubule as the locus, in a person's avatar, of the physical correlate of consciousness—specifically, the correlate of the portal between the personal mind and the metamind. It is known that the cytoskeleton of the microtubule can sustain cellular automata [21]. Therefore one avenue of inquiry for narrowing down the space of possible of cellular automata with the framework given above, would be to examine mental cellular automata that could be mapped onto cytoskeletal cellular automata.

4. Conclusions

Despite the solid philosophical arguments that have been advanced for mental monism [4,10], this philosophical doctrine often encounters a visceral opposition. In part this seems to be due to unconsidered adherence to the metaphysical hypothesis of physical reality, which renders mental monism too outrageous to contemplate. It is also due to the absence of any workable model for the structure and dynamics of not only the conscious mind but the whole manifold of observed regularity that we have successfully modelled with physical laws over the past three centuries, which mental monism must ultimately re-ground in the mind.

What the arguments above show is that, with parsimonious and philosophically plausible hypotheses we can formulate a class of formal models of the conscious mind, some of which can sustain universal computation in the form of cellular automata. The arguments above formulate a tentative basic model. This conclusion points to a two-pronged research programme: on the one hand computer simulations of the relevant kind of cellular automata; on the other hand investigation of physical correlates of consciousness that exhibit behaviour akin to cellular automata. The goal is, obviously, a long way off.

Funding: This research received no external funding.

Acknowledgments: Thanks to the organisers of the conference *Models of Consciousness: A conference on formal approaches to the mind-matter relation* for allowing me to give a brief presentation of this material [22].

Conflicts of Interest: The author declares no conflict of interest.

References

1. Descartes, R. *Meditations on First Philosophy, in which the existence of God and the immortality of the soul are demonstrated*. Originally published 1641. Veitch, J., ed.; Dent: London, 1975.
2. Chalmers, D. *The Conscious Mind*. OUP: Oxford, 1996.
3. Lloyd, P.B. Mental Monism Considered as a Solution to the Mind-Body Problem, in Batthyany, A. & Elitzur, A., eds., *Mind and its Place in the World: Non-Reductionist Approaches to the Ontology of Consciousness*, Ontos Press: Frankfurt, 2007; pp. 101-144. Available online: http://www.peterblloyd.com/consciousness/Mental Monism Considered as a Solution to the Mind-Body Problem.pdf
4. Lloyd, P.B. Panpsychism and Mental Monism: Comparison and Evaluation. 2019. Available online: https://www.researchgate.net/publication/332978948_Panpsychism_and_Mental_Monism_Comparison_and_Evaluation DOI: 10.13140/RG.2.2.30580.60806.
5. Hameroff, S. & Penrose, R. (2014). Consciousness in the universe: A review of the 'Orch OR' theory, *Physics of Life Reviews* 11, pp. 39–78. https://doi.org/10.1016/j.plrev.2013.08.002
6. Berkeley, G. *The Principles of Human Knowledge*, Volume I. James Pepyat: Dublin, 1710.
7. Foster, J. *The Case for Idealism*, Routledge & Kegan Paul: London, 1982.
8. Lloyd, P.B. *Consciousness and Berkeley's Metaphysics*, 281 pp. Self-published: London, 1999.
9. Lloyd, P.B. *Paranormal Phenomena and Berkeley's Metaphysics*, 353 pp. Self-published: London, 1999.
10. Pearce, K.L. *Language and the Structure of Berkeley's World*, PhD thesis, University of Southern California, 2014. http://writings.kennypearce.net/diss.pdf. Published as: Language and the Structure of Berkeley's World, OUP: Oxford, 2017.
11. Haggard, P.; Boer,L. Oral somatosensory awareness. *Neuroscience & Biobehavioral Reviews* **2014**, *47*, 469-484. Available online: https://doi.org/10.1016/j.neubiorev.2014.09.015.
12. Caputo, G.B. Strange-face-in-the-mirror illusion. *Perception* 2010, **39**, pp 1007-1008. DOI: 10.1068/p6466 Available online: https://www.researchgate.net/publication/46280355_Strange-face-in-the-mirror_illusion
13. Block, N. (1978). Troubles with Functionalism, in Savage, C.W. (ed.), *Perception and Cognition: Issues in the Foundations of Psychology*, Minneapolis: University of Minnesota Press.
14. Gardner, M. Mathematical Games – The fantastic combinations of John Conway's new solitaire game 'life'. *Scientific American* 223 (4): 120–123. October 1970. doi:10.1038/scientificamerican1070-120.
15. Cook, M. Universality in Elementary Cellular Automata. *Complex Systems* 15, 1-40, 2004.
16. Hoffman, D. *The Case Against Reality*. Harmondsworth: Penguin Books Ltd. Kindle Edition.
17. Hoffman, Donald D. (Sensory Experiences as Cryptic Symbols of a Multimodal User Interface. *Activitas Nervosa Superior* **2010**, 52 (3): 95–104. doi:10.1007/BF03379572.
18. Kastrup, B. An ontological solution to the mind–body problem. *Philosophies* **2017** 2:2.
19. Kastrup, B. The universe in consciousness, *Journal of Consciousness Studies* **2018** 25 (5-6) pp. 125-155.
20. Kastrup, B. & Kelly, E.F. Could Multiple Personality Disorder Explain Life, the Universe and Everything?, *Scientific American* blog, 18th June 2018. https://blogs.scientificamerican.com/¬observations/could-multiple-personality-disorder-explain-life-the-universe-and-everything/
21. Smith, S.A., Watt, R.C., Hameroff, S. Cellular automata in cytoskeletal lattices. *Physica D: Nonlinear Phenomena* 1984, **10**(1-2), pp 168-174.
22. Lloyd, P.B. Automata-theoretic approach to modelling consciousness within mental monism. Presented at *Models of Consciousness: A conference on formal approaches to the mind-matter relation.* Mathematical Institute, University of Oxford, September 9 - 12, 2019. http://podcasts.ox.ac.uk/peter-lloyd-automata-theoretic-approach-modelling-consciousness-within-mental-monism