

# Deconstructing the Physical World

Brendon Hammer

**Abstract:** Some metaphysics are provided showing that what is commonly called ‘the physical world’ can be deconstructed into three ‘levels’: a single, unified ‘noumenal world’ on which everything supervenes; a ‘phenomenal world’ that we each privately experience through direct perception of phenomena; and a ‘collective world’ that people in any given ‘language using group’ experience through learning, using and adapting that group’s language. This deconstruction is shown to enable a clear account of qualia and of how people can hold some things to be physically real even when it is clear no one can ever directly experience those things as phenomena. It is further shown to enable a single, internally consistent, largely empirically supported conceptual framework – a ‘metacosmology’ – able to encompass not only the physical world as people conceive of it for everyday purposes, and as scientists conceive of it for scientific purposes, but also people’s first person phenomenal experience of a physical world and, prospectively, the mechanisms by which such first person, conscious experiences can be generated.

## 1. Phenomenal worlds and the idea of a noumenal world

First consider a person’s moment-by-moment phenomenal experience of themselves as a physical being – a physical body – perceiving and interacting with a physical world.

From at least the time of Kant (1) various schools of thought have held that a person’s phenomenal experience of the physical world must arise through some form of processing of sensory input within the person and – if their intentional interactions with the world are taken into account – the genesis and expression of motor output into that world through further processes within the person.

A subsequent emerging mainstream view – along with a range of hypotheses – has been that both these processes occur within a person’s body, largely or wholly within their brain (2). Important contemporary applications of this view can be found, for example, at the core of the phenomenal self-model (3, 4, 6) and predicative processing (8, 10) paradigms, where the latter explicitly envisages a brain-based processor that integrates implementation of both processes (8, 10, 13, 14). For the purposes of this note assume then that a processor exists – call it D – that implements both processes, and that D is part of a person’s brain.

A further widely held assumption is that the source of a person’s sensory input and the destination of their motor output is a single physical world that all human beings inhabit, observe/perceive and interact with, that contains all physical things, but whose existence, contents and behaviour are negligibly affected by whether or not it is being observed/perceived – i.e. experienced as phenomena – by any person.<sup>1</sup> The practice and theory of virtually all mainstream physical science assumes such observer independence – which is that, absent any motor or other output from them into it, any person’s observation/perception of a physical system – i.e. their phenomenal experience of it – has negligible effect on that system’s contents or transformations.<sup>2</sup>

So for the purposes of this note a core materialist assumption will be adopted that there exists a single physical world, containing all physical things, that is the source of all information capable of sensory perception, regardless of whether any observer is present to receive and process such information to form perceptions. This world will also be the destination for all motor output generated by a person.

Call this assumed single<sup>3</sup> physical world the *noumenal world*,  $W[r]$ , where  $W[r]$  can be considered the world as it exists independently of acts or experiences of perception, where all information capable of sensory perception originates, and upon which all human motor output operates.<sup>4</sup>

Hence  $W[r]$  can be defined as the set of all things<sup>5</sup> that exist independently of whether or not they are under perception as phenomena, and which includes all sources of information capable of sensory perception. On this basis, earmark any given thing,  $T_n$ , that is a member of  $W[r]$  by giving it the suffix  $[r]$ , where it can be put that:

$$W[r] = \{T_1[r], T_2[r], T_3[r], \dots, T_n[r], \dots\} \quad E1$$

<sup>1</sup> Issues of observer independence, including debate about the Copenhagen interpretation of quantum mechanics, apply to events at scales in space and time that are far too small to be relevant to the discussion here, which deals only with the macroscopic scale of everyday human phenomenal experience of the physical world and the activities and results of science manifest at that scale. David Wallace provides authoritative contemporary support for observer independence (15), which was also considered to apply by a key architect of the Copenhagen interpretation, Niels Bohr (16).

<sup>2</sup> Where ‘transformations’ mean changes of any kind, i.e. ‘motions’ of a physical system in the broadest sense.

<sup>3</sup> Where by *single* physical world is meant a physical world of things unified through each having relations to all other such things under one logically coherent set of rules. This aligns with a metaphysics where the universe “is fundamentally a unitarily evolving whole” as proposed, for example, by Wallace (17).

<sup>4</sup> NB: For the purposes of this note the terms ‘noumenal world’ and ‘phenomenal world’ are meant *only* as they are defined *within this note*, and as above. For example, what is here being defined as the noumenal world is defined such that it is at least partially, if *indirectly*, knowable.

<sup>5</sup> Here the term ‘things’ is a placeholder where, under *any* parsing of the contents of the set  $W[r]$  into members, such members can be called ‘things’.

Now define a *phenomenal world*, call it  $W[i]$ , as being the set of all things – including all relationships, motions<sup>6</sup> and properties of those things – that constitute all of the moment-by-moment contents of a person’s phenomenal experience of themselves as a physical body operating within, and interacting with, a physical world. Earmark any given thing,  $T_n$ , that is a member of a  $W[i]$  with the suffix  $[i]$ , where it can be put that:

$$W[i] = \{T_1[i], T_2[i], T_3[i], \dots, T_n[i]\} \quad E2$$

We each have direct familiarity with the contents of our individual phenomenal worlds, including our phenomenal body, and we are used to perceiving such things as a table $[i]$ , a forest $[i]$ , a tree $[i]$ , a leaf $[i]$  or a river $[i]$ , where each has by some means been parsed such that we perceive these as more-or-less discrete members of our  $W[i]$ .

Now consider that if:

- (a) phenomenal experience arises for a person through processing in  $D$  of sensory information originating in  $W[r]$

and

- (b) some kind of rules-based connection (RBC)<sup>7</sup> between the contents of  $W[r]$  and the contents of their  $W[i]$  is implicit in such processing

then

- (c) for any given moment of a person’s phenomenal experience of their  $W[i]$ , an inversion of whatever RBC gives the specific contents of that experience will *map* the things that make up those contents back onto, and thereby define,<sup>8</sup> some specific contents of  $W[r]$

such that

- (d) for any given moment of a person’s phenomenal experience it will be possible to take any thing that the RBC has parsed to be a member of their  $W[i]$  – say  $T_1[i]$  – and to further define a thing,  $T_1[r]$ , as being that member of  $W[r]$  which the inverted RBC defines as producing the information that, when processed by  $D$ , gives a phenomenal experience of the thing,  $T_1[i]$

where

- (e) a way that  $W[r]$  can specifically be parsed into members,  $T_1[r], T_2[r], T_3[r] \dots T_n[r]$ , can then be defined by the relationships specified in (d).<sup>9</sup>

Using this approach, the assumption of an RBC makes it possible to populate  $W[r]$  with things that can be defined as those sources (hidden causes) of information that determine, through that RBC, the things that make up the contents of a person’s phenomenal experience of their  $W[i]$ .

It can then be maintained that any phenomenal thing  $T_1[i]$  – say a tree $[i]$  or leaf $[i]$  – that a person experiences as a member of their  $W[i]$  will have a *correlate* noumenal thing  $T_1[r]$  – a tree $[r]$  or leaf $[r]$  – that is a member of  $W[r]$ , recalling that by the earmarking convention proposed above, any given thing,  $T_n$ , that is a member of  $W[r]$  is to be given the suffix  $[r]$ .<sup>10</sup> This approach allows for more precise definition of some key terms used earlier; specifically: ‘physical body’, ‘brain’ and ‘processor  $D$ ’.

<sup>6</sup> Where the meaning of ‘motions’ is as provided in footnote 2, and will include phenomenal experience of relative motions of object $[i]$ s but also, for example, of changes in the shape of object $[i]$ s, in temperature $[i]$ , in illumination $[i]$  and so forth.

<sup>7</sup> For now no attempt will be made to further describe an RBC (apart from the working assumption made at footnote 10). It is simply proposed, given earlier assumptions, that for any given moment of normal waking phenomenal experience such a connection will exist and operate. Notably however, predictive processing approaches (7, 9, 12, 13) make detailed proposals about RBCs and how they might operate, including proposing that any given person’s RBC will evolve over time. The possibility of such evolution is taken into account here by applying (a) to (d) to situations caveated to be at, ‘any given moment’. This caveat will nevertheless be set aside for the remainder of this note under an assumption that any evolution of a specific person’s RBC will, for a typical adult under typical circumstances, be sufficiently slow and small that their RBC can be assumed to be unchanging with respect to the ideas to be presented in the note.

<sup>8</sup> Where it can be assumed that such definition will be to within some margin of error, and may to some significant degree be probabilistic.

<sup>9</sup> By which means the placeholder ‘things’ and the parsing referred to in footnote 5, is now more clearly defined.

<sup>10</sup> (a) Note that any thing $[i]$  that is a member of a  $W[i]$  might equally be considered the name of a set that is a subset of that  $W[i]$ , where such a subset may itself have thing $[i]$ s as its members and so on down to smallest perceivable units. For example, the term leaf $[i]$  can be used to name a thing $[i]$  that is a member of a subset of a  $W[i]$ , where that subset can itself be a thing $[i]$  named a tree $[i]$ , which may itself be considered a member of a subset which is a thing $[i]$  named a forest $[i]$ . This is empirically grounded in our phenomenal experience that a given thing $[i]$  is often made up of other thing $[i]$ s, where those other thing $[i]$ s are perceived to form ‘parts’ of that given thing $[i]$ . (b) With this in mind, assume for now that for any given moment the way all actual RBCs map members and subsets of  $W[r]$  into a  $W[i]$  and, by inversion, map members and subsets of a  $W[i]$  into  $W[r]$ , conserves the order in which things form parts of other things. For example, it will be assumed that in the case where leaf $[i]$  is a member of a set tree $[i]$ , leaf $[r]$  will be a member of a set tree $[r]$ , and so on for all cases such that where any  $a[i]$  is a member of the set  $b[i]$  then whatever RBC applies will give  $a[r]$  as a member of the set  $b[r]$ . Note further then, that being a member of any such set will from hereon also be referred to as being ‘in’, ‘within’, ‘inside’ or ‘part of’ that set. These terms help in defining the relative locations of the sources or destinations of information flow, and of where information is being processed; for example flows of sensory and motor information as illustrated in Figure 1 (below).

So where a person's *phenomenal experience* of themselves as a 'physical body' – as it operates within and interacts with a 'physical world' – has previously been referred to, this can now more precisely be referred to as a person's experience of themselves as a:

- 'physical body[i]' – from hereon called their *phenomenal body*, denote this B[i] – which is their physical body as they perceive it operating within, and interacting with
- their 'physical world[i]' – that is, their *phenomenal world*, W[i] – in terms of their perception of their B[i] operating, and interacting with other thing[i]s, within their W[i].

Although people rarely perceive their own brain as a physical phenomenon, it is not hard to imagine how this could be achieved using surgery and mirrors. In such a scenario what a person would perceive would be their brain[i].

It can now be proposed, through consideration of (a) to (e) above, that for any person it can be inferred from their personal phenomenal experience that there exist such correlate noumenal thing[r]s as their brain[r], their *noumenal body*, B[r] and, as previously described, a noumenal world, W[r], that will be experienced by that person as their phenomenal world, W[i].

Moreover, *if* – consistent with the orthodoxy of mainstream cognitive science – a person's experience of a phenomenal world and its contents is generated through processing of sensory input by some processor D within that person's brain, these *must* be references to a proposed processor D[r] within that person's brain[r].<sup>11</sup>

Figure 1 schematically describes the structures and information processing arrangements proposed above.

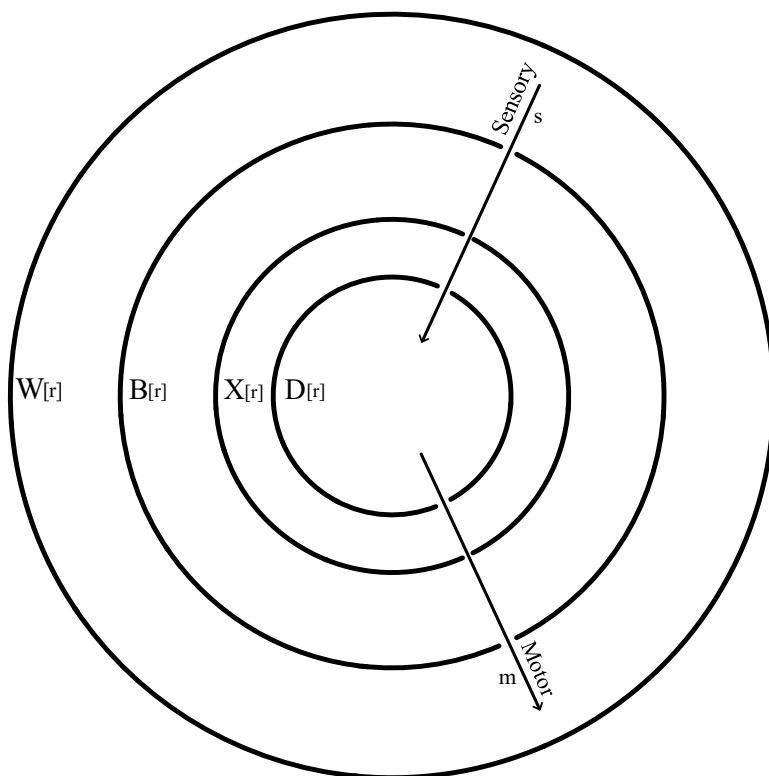


Figure 1

**Figure 1** shows the structures and information processing arrangements proposed in the text. W[r] is the noumenal world, defined as the set of all thing[r]s that exist independently of whether they are experienced as phenomena by any person. W[r] will be the source of all sensory input, s, into the proposed processor D[r], and the destination of all motor output, m, from D[r]. B[r] is the noumenal body, which is the subset of W[r] that contains all thing[r]s that make up a person's physical body[r]. X[r] denotes a person's nervous system[r] including their brain[r], within which their D[r] is housed. Overall, the figure can be considered a Venn diagram of a steady state structure  $D[r] \subset X[r] \subset B[r] \subset W[r]$  within which the information processing arrangements proposed in the text take place, consistent with the assumptions and definitions given at footnote 10(b).

<sup>11</sup> This is simply since the idea of information being processed within someone's W[i] – i.e. within their *perceptions* of the physical world – makes no sense, whereas the idea of information being processed within W[r] makes clear sense.

## 2. Alignment of Phenomenal Worlds

Under the arrangements proposed above each person will at any given moment directly experience their own discrete phenomenal world, their  $W[i]$ , generated by and within their own  $D[r]$  as it applies its own RBC to sensory input from the noumenal world,  $W[r]$ .

Consistent with this, it is empirically clear to each of us that we cannot experience any other person's phenomenal world as they directly subjectively experience it.

Yet we also know that the contents of other people's phenomenal worlds contain a vast number of features that have *something* in common with what we ourselves experience in our own phenomenal world, and that a great many of those *common features* prove empirically to be ubiquitous and enduring.

We cannot be born knowing which contents of our phenomenal experience we will be able to hold to be common features, since at time of birth we will not have had sufficient interaction with others – or with  $W[r]$  – to have gained such knowledge. But without need for reflection, from infancy, human beings naturally and rapidly learn to identify and verify an immense number and variety of these common features through learning and use of language.<sup>12</sup> This is how our species discovers and can then exploit such common features: by naming them, describing them and communicating about them in ways that facilitate cooperative physical action – through motor output from our bodies – to organise and alter such features to our benefit.

Here then, when people – including scientists and engineers – refer to “the physical world”, what we are referring to in *all* of our communications with each other can *only* be those things that we collectively agree to be common features among what we find in our respective  $W[i]$ s.<sup>13</sup> As individuals we find and agree on these features qualitatively through learning and using language, and quantitatively – for example, in science and engineering – by augmenting language through learning and using agreed measuring procedures.

Learning to identify, name and communicate about such common features allows each of us to operate on the basis of a working assumption so natural that we never question it in everyday life. This assumption is that our phenomenal world – that is, our  $W[i]$ , including our  $B[i]$  – is a veridical experience of a single physical world; a world that we share with all other people and with all other physical things. This is despite the fact that the only world any of us will ever be able directly perceptually to experience is our own discrete and, in this sense, unique  $W[i]$ .

This assumed single physical world is what people routinely refer to as “the physical world”. It is a *presumed* world that we collectively and *tacitly* agree – through learning and use of language and agreed measuring procedures – to be that which constitutes *the overall ensemble of common features* that we each can identify across our respective  $W[i]$ s.

To restate and reemphasise this point, “the physical world” as we each refer to it in *all* of our communications, will actually be a ‘world’ – an *abstraction* from hereon denoted  $W[z]$  – made up only of that which we each tacitly agree with other members of our *language using group*<sup>14</sup> are common features of our respective  $W[i]$ s, based on our collective identification and agreement on those common features through our learning and use of language and agreed measuring procedures.<sup>15,16</sup> With this understood, the abstraction  $W[z]$  can be considered to be the collectively agreed physical world or, in shorthand, the *collective world*, as each of us holds it to be.<sup>17</sup>

So far then, three ‘versions’ – call these *levels* – of ‘physical world’ have been described.  $W[r]$ , the noumenal world as defined above;  $W[i]$ , an individual person's phenomenal world – i.e. the physical world as each person individually and *directly* phenomenally experiences it; and the abstraction  $W[z]$ , which is an *assumed* collective world that we each tacitly take to constitute a single *shared* physical reality, based on a collective conceptual *alignment* of our individual phenomenal worlds achieved through our individual learning and use of language and of agreed measuring procedures.<sup>18</sup>

<sup>12</sup> What is meant here by ‘language’ is *any* form of interpersonal communication intelligible to the language using parties.

<sup>13</sup> There is more to it – see *imaginary common features* (footnote 16 and Section 5(ii) p9) – but this formulation has sufficient strength to serve for now.

<sup>14</sup> For the purposes of this note it will from now on – unless otherwise specified – be assumed that the extent and modes of communication among people currently is such that – despite “language barriers” – we can more-or-less all be considered members of the same language using group. In particular, this claim holds for the international community of scientists in relation to their common practice and communications about science and for all science-relevant aspects of the  $W[z]$  they share.

<sup>15</sup> In other words, we build and maintain a collectively held abstraction, a  $W[z]$ , by each holding within it all the common features we are each able to identify in our respective  $W[i]$ s by naming and describing them as we individually go about learning, using and adapting our respective languages and agreed measuring procedures. We do this without reflection because, like learning to walk, it is an entirely natural process for our species.

<sup>16</sup> Importantly, a person's  $W[i]$  may sometimes permanently lack – and they will be unable phenomenally to experience – one or more of the ubiquitous and enduring common features that others collectively can identify through language learning and use as present in each of their  $W[i]$ s. For example, people with red-green colour blindness can never directly experience, or name from direct experience, red as distinct from green. In such cases the person must then infer, by taking others' ‘word for it’, that the given common feature exists and is to be placed in  $W[z]$  – e.g. that distinct colours red and green exist and should be placed in  $W[z]$  – just as if it were possible for them directly to perceive those colours. This ability to identify, name and place common features into  $W[z]$  *purely* on the basis of participating in language use about them, although such common features may never be available for direct perception in a person's  $W[i]$ , is important in discussion below of *imaginary common features*. Atoms are an example of imaginary common features. Names and descriptions of imaginary common features can be part of the contents of  $W[z]$  while at the same time those common features may have no availability for direct perception in anyone's  $W[i]$ .

<sup>17</sup> NB: Because of how it is learnt, each person's knowledge of  $W[z]$  – i.e. what they hold to be  $W[z]$  – can be considered equivalent to their overall conscious and unconscious knowledge of language as it relates to their identifying, naming, describing, measuring and manipulating contents of the physical world.

<sup>18</sup> The work here was developed independently of Karl Popper's ideas and has no reliance on them. Although the ‘levels’ of physical world proposed above overlap conceptually with Popper's ‘world 1’, ‘world 2’ and ‘world 3’ (18, 19) there are crucial differences. Most importantly, it is proposed here (Section 6) that *all* contents of any person's  $W[i]$  or  $W[z]$  will supervene on contents and information processing (changes of state) in  $W[r]$ , and can have no ‘reality’ *independent* of that; i.e. they can have no ‘stand alone’ – objective – reality. One of Popper's central claims is that his worlds 1, 2 and 3 each have objective reality (19).

### 3. The Three World Praxis

For everyday purposes we each operate on a compound and *tacit* assumption – essentially, a praxis – that W[r], our individual W[i]s and our W[z] are identical to, and can be held to be in identity with, each other. That is, we operate on a tacit working assumption that no distinction need be drawn between W[r], our W[i]s and our W[z], and that these can all be held to be the same thing.

In maintaining this praxis we each, for all practical purposes, tacitly assume – and act – as if:

- α) the physical world as we each experience it phenomenally, in the form of our W[i], is at any given moment a direct veridical experience of – and may be held to be a world identical to, and in identity with – part of a single unified physical world that exists independently of all of us, at the noumenal level, i.e. W[r]; and that
- β) the physical world as we each experience it phenomenally, in the form of our W[i], presents itself in exactly the same way phenomenally – and may be held to be a world identical to, and in identity with – the physical world as it is experienced phenomenally by all other people in the form of their respective W[i]s.<sup>19</sup>

Call this compound assumption – (α) with (β) – the *three world praxis* (TWP).

Notably, we do not apply assumption (β) so rigorously that we expect other people to have utterly the same phenomenal experience of the world that we do. For example, we accept that at any given time each person will experience the world from a different vantage point; that is, from their own unique spatial perspective, be that near or far from our own spatial perspective. But beyond this we do routinely assume that when, for example, we and another person both see a specific red, that person is having exactly the same direct perceptual experience of seeing that red that we are having, or that when we point to a chair and say it is broken that they are seeing the same chair, and experiencing it phenomenally in exactly the same way that we do, aside from spatial perspective.

We apply the TWP because it turns out that we *can* and, more importantly, because its application works incredibly well for us. Indeed, we are able constantly to reinforce our tacit presumption that the TWP is sound – along with our *sense* that our presumed collective world, W[z], is the same single shared ‘physical reality’ – through our ongoing *successful* use of language to agree, either explicitly or implicitly, on the common features in our respective W[i]s and – *crucially* – through our use of language to plan and then to implement agreed cooperative physical actions that *successfully* modify and test the physical world in specific and *jointly agreed* ways.

A zenith here is our cooperative design and use of tools, including machines and habitats, to extend such success in ways that give us individually, and as a species, a mastery of our physical environment and our other physical circumstances that is utterly unmatched by any other animal.<sup>20</sup>

Yet if it is to be claimed that the TWP is a working assumption, not a reality, and that W[r], our W[i]s and W[z] are not identical to each other – let alone in identity with each other – it becomes important explicitly to answer the questions:

- 1) How is it that we can apply the TWP so successfully?
- 2) In what ways do W[r], W[i] and W[z] differ from each other?

### 4. How is it that we can apply the TWP so successfully?

The answer to this question rests on three requirements. The first (i) is that for the three world praxis (TWP) to work, there must really be a single unified noumenal world, W[r], within which all people exist at the level of their respective B[r]s, which is the source of all sensory input that their individual D[r]s receive and process through their individual RBCs to give their individual W[i]s, and which is the destination of all of their motor output.<sup>21</sup>

---

<sup>19</sup> Hence, and to put (α) and (β) into more compressed but familiar terms: We each assume that every person’s W[i] – including our own – provides us each as individuals with a direct, veridical experience of a single, unified physical world that exists independently of all of us.

<sup>20</sup> Given that no other animal species has the extent of language use that we do then, under the arrangements described above, no other species will be able to establish to anything like the extent that we do a shared sense among its members that they are all inhabiting the same, single physical world.

<sup>21</sup> This restates the core materialist assumption made in Section 1 of this note, where – with specific regard to (i) above – it is clear that if this assumption were incorrect our respective physical realities would lack common ground. In other words, if we did not all as B[r]s inhabit a single unified noumenal world then the foundations for our development of a consistent common agreement on what constitutes our shared physical environment would have to be flawed. Depending on the extent of this flaw, such common agreement – i.e. the shared abstraction W[z], as held by each of us – would either not exist or should break down at some point. This is not what we observe, including at the limits of our understanding of physics, which can be considered the cutting edge of our development of W[z] (see Section 5(iii) below), and where so far, to the best of our knowledge, all scientists reproducibly observe the same results for any given specific experiment.

Second (ii), in addition to (i) the RBC and the D[r] information processing system we each have – though unique to each of us – must be sufficiently *similar* to those of other people to be able to deliver to us each a direct experience of a W[i] that exhibits phenomena such that we can each then successfully identify and verify, through language learning and use, a range of common features that we can agree those phenomena share with phenomena other people directly experience in their W[i]s.

Third (iii), it must be the case that the common features we are each able to identify through learning and use of language – and can then seek to transform through agreed collective action to achieve agreed outcomes – are such that they will always, at least for all practical purposes, *correspond* to things in W[r] that can be transformed at the level of W[r] by motor output from our B[r]s into W[r] such that an agreed outcome at the level of said common features – i.e. at the collectively acknowledged level of W[z] – can actually successfully be achieved.

Requirement (i) is a matter of ontology. There is either a single unified noumenal world, W[r], or there is not. Our experience of successful human application of the TWP provides strong empirical evidence that W[r] really does exist.<sup>22</sup>

Requirements (ii) and (iii) – which must also be met for the TWP successfully to be applied – depend on how each person’s RBC and D[r] processors are designed and operate.

Specifically, for the TWP to be applied successfully, people’s individual RBCs and D[r]s must not only be sufficiently similar to allow identification by each person – through language learning and use – of common features in their respective W[i]s, but this similarity must be convergent on a common brain[r] information processing architecture that ensures each person’s direct phenomenal experience of those common features is such that those common features *correspond* to things in W[r] as required under (iii).

These observations in relation to requirements (i) to (iii) provide for an answer to question 3(1): How is it that the TWP can be applied so successfully?

First – with these observations in mind – and on empirical grounds of our highly successful experience of implementing the TWP – it can be sustained that W[r] must indeed exist, and that W[r]’s existence *must be such that* information processing systems can arise within it that meet requirements (ii) and (iii) above. It can then be proposed that the *immense survival advantage* conferred upon our species through our ability to apply the TWP – as outlined in Section 3 above – will have driven evolution of the human brain[r] and body[r] such that all species-typical human beings develop D[r]s and RBCs that meet requirements (ii) and (iii).

The claim here then is that Darwinian selective pressures will have driven, and will keep driving, evolutionary convergence onto a human physical[r] and neurological[r] architecture that maximises the similarities required under (ii) and (iii) among the D[r]s and RBCs of each person, such that – as each person individually develops and grows to learn about their individual W[i]s – they are also able to learn and use language to develop a sense of W[z], and to then implement the TWP to massively enhance, in cooperation with other people, both their own and the overall human species’ potential for survival as physical beings in a physical world.<sup>23</sup>

## **5. In what ways do W[r], W[i] and W[z] differ from each other?**

### *(i) Differences between W[i] and W[r]*

One example of where a person’s individual W[i] can differ from W[r] arises when what they directly perceive is mistaken. For example, in low light it is quite possible that a person might for some moments see, in a particular play of light and shadow among a pattern of stems and leaves, a man standing some distance away at the edge of a garden. This man could seem perfectly real right up until a gust of wind disturbs the relevant stems and leaves and the impression is shattered. There never was such a man, but for a short period there genuinely was a man in the person’s W[i], making it different – here, mistaken – in its connection (correlation) to what was in W[r]. Military use of decoys and of camouflage exploit people’s propensity to be mistaken in what they directly perceive in their W[i].

---

<sup>22</sup> As per the opening paragraph of Section 4 and, in particular, footnote 21.

<sup>23</sup> In this respect collective human application of the TWP can be considered an ongoing, adaptive enterprise that massively enhances the Darwinian fitness of the human species.

Such mistakes fall into a general category of *perceptual illusions*. A perceptual illusion is what we call something that we directly perceive in our  $W[i]$  that can – by further perceptual exploration or by other means – be determined to be, or to have been, erroneous in its mapping of the hidden causes in  $W[r]$  that have given rise to it.

A related, far starker difference between what a person experiences in their  $W[i]$  and what is underlying that experience in  $W[r]$  arises when they are asleep and dreaming. No one finds that what happens in the  $W[i]$  they experience while dreaming – although it may seem quite real as they directly perceive it during dreaming – is what has actually happened in the single physical world they hold themselves to inhabit while awake. It turns out that in a dream what a person directly perceives to be in their  $W[i]$  is radically different – here, radically ‘mistaken’ – in relation to what can have been underlying those direct perceptions in  $W[r]$ .<sup>24</sup>

### (ii) Differences between $W[i]$ and $W[z]$

To identify differences between the contents of any given person’s  $W[i]$  and the contents of what they hold to be the collective world  $W[z]$ , it is useful to recall that:

- unlike any of our individual  $W[i]$ s, which we clearly and directly perceive as an integrated and continuous physical world surrounding us and containing our physical body; and
- unlike  $W[r]$  which, although we do not experience it directly, we readily can conceive of as being a single unified world that exists in its own right<sup>25</sup>

the  $W[z]$  that we each maintain cannot be considered to have any existence independent of the language and agreed measuring procedures that we learn to use to describe its contents.<sup>26</sup>

As such, that which we each learn to hold and maintain as  $W[z]$  can only be considered a ‘world’ to the extent of its being an abstraction formed purely by our each finding, identifying and then holding within it as its contents that which we learn – by *successfully* learning to name, describe, measure and *collectively manipulate* them – are common features that we can each *also* find and identify within our own  $W[i]$ s (or that we can infer to be as good as found and identified within our own  $W[i]$ s<sup>27</sup>).

---

<sup>24</sup> One possible explanation for dream  $W[i]$ s is that – consistent with the ideas developed in Section 1 – a dreaming person’s  $D[r]$  is applying the same RBC it uses when they are awake to input that is *not* sensory input from outside their  $B[r]$ , such that their  $D[r]$  generates a  $W[i]$  that contains phenomenal things that connect to what is happening in  $W[r]$  in an entirely different way to the way their  $W[i]$  is connected to  $W[r]$  when they are awake. More specifically, in dreaming one possibility as to what could be happening in  $W[r]$  is that some kind of signal is being generated outside the person’s  $D[r]$ , yet still inside their brain $[r]$ , and is entering their  $D[r]$  through the same pathways that usually pass sensory information into  $D[r]$  when they are awake, to then be processed through their RBC to evoke the strange  $W[i]$  of the dreamer.

A similar explanation for perceptual illusions can be made based on a proposal already implicit in Section 1 that, *in the ideal case*, all of the time a person is awake there will be a fully stable, entirely consistent processing of sensory input by their  $D[r]$  through application of a fully stable, entirely consistent RBC such that the person will *always* directly perceive at the level of their  $W[i]$  some given thing as  $T_1[i]$  when the source in  $W[r]$  of the sensory input being so processed is  $T_1[r]$ .

Some illusions could then arise in certain kinds of non-ideal case where, due to an inconsistency in the processing of sensory input by a person’s  $D[r]$  the person directly perceives some given thing  $T_6[i]$  – instead of directly perceiving a different given thing,  $T_2[i]$  – when the source in  $W[r]$  of the sensory input being so processed should, if operation of their RBC was ideal, actually be  $T_2[r]$ . So here an illusion can be considered a processing error relative to the ideal case, where in the ideal case *all* examples of direct perception of any thing  $T_n[i]$  the source in  $W[r]$  of the sensory input being processed should always be  $T_n[r]$ , and never  $T_{(n+1)}[r]$  or any other thing besides  $T_n[r]$ . So for example, in the case of the illusory man at the garden’s edge, some kind of *non-ideality* – a limit or flaw – in the person’s  $D[r]$  processing capability has given them a direct perception of a man =  $T_6[i]$ , and not a direct perception of a play of light and shadow in some stems and leaves =  $T_2[i]$ , when the source in  $W[r]$  of the sensory input being processed has actually been  $T_2[r]$ . In this example, in due course – with shattering of the illusion – what the person directly perceives does then reveal itself to be  $T_2[i]$ .

Dreams and many types of perceptual illusion can be considered normal. *Hallucinations*, particularly visual hallucinations where a waking person directly perceives in their  $W[i]$  physical things that by some other means can be shown to be impossible, are relatively rare and tend to indicate some form of neurological damage or abnormality. Consistent with the explanations proposed above for illusions and dreaming, an explanation for a person’s hallucinatory experience of physical things – i.e. of seeing things that ‘aren’t there’ – seems likely to be that during such hallucinations some serious malfunction is occurring in that person’s  $D[r]$ ’s ability to sustain and properly implement an RBC.

<sup>25</sup> And which, as described in Section 1, gives rise to the information – and contains and implements those information processing arrangements, including each of our  $D[r]$ s and RBCs – by which our brain $[r]$ s generate our individual  $W[i]$ s.

<sup>26</sup> In this sense  $W[z]$  can be thought of as a ‘world’ ‘made of language’; i.e. a ‘world’ that a person can *only* build – and come to participate in – by learning to communicate with others (even though – as members of the human species – people go about this learning, building and participation with the same kind of non-reflective automaticity with which they learn to walk).

<sup>27</sup> See footnote 16 for an explanation of what is meant here by such inference and note the concept of ‘imaginary common features’ that will be introduced below (Section 5(ii) pp9).

If this is so, then finding a way to describe by use of language that which, by definition, should be ‘indescribable’ by use of language – i.e. any contents of a person’s  $W[i]$  that *by definition* cannot be found, identified and held within what they hold to be  $W[z]$  – may seem impossible. But if such contents *could* be found in a person’s  $W[i]$ , they would constitute a difference between the contents of their  $W[i]$  and anything that they could ever find, identify and hold to be contents of  $W[z]$  through learning and use of language.

As it happens, it is possible to demonstrate that contents of this kind do exist without actually having to describe them. Instead, it is possible to describe a ‘cognitive procedure’ by which a person can privately conceptually isolate such contents. When multiple people are able successfully to use such a procedure to conceptually isolate such contents they can then agree on a name for the kind of contents they each understand themselves to have isolated, even though they may share no other means of identifying or describing those contents beyond pointing to the use of such a procedure.

An example of a procedure of this kind is provided by the thought experiment called Mary’s Room (20(a), 20(b)). A version of this experiment, the same as that described by Michael Tye (20(b)), is as follows:

Mary is imprisoned in a black and white room. Never having been permitted to leave it, she acquires information about the world outside from the black and white books her captors have made available to her, from the black and white television sets attached to external cameras, and from the black and white monitor screens hooked up to banks of computers.

As time passes, Mary acquires more and more information about the physical aspects of colour and colour vision. Eventually, Mary becomes the world’s leading authority on these matters. Indeed she comes to know *all* the physical facts pertinent to everyday colours and colour vision.

Still, she wonders to herself: What do people in the outside world *experience* when they see the various colours? What is it *like* for them to see red or green? One day her captors release her. She is free at last to see things with their real colours (and to scrub off the black and white paint that covers her body). She steps outside her room into a garden full of flowers. “So, that is what it is like to experience red,” she exclaims, as she sees a red rose. “And that,” she adds, looking down at the grass, “is what it is like to experience green.”

Mary here seems to make some important discoveries. She seems to find out things she did not know before. How can that be, if, as seems possible, at least in principle, she has all the physical information there is to have about colour and colour vision — if she knows all the pertinent physical facts?

In the context of the ideas being described in this note, what Mary has discovered – and what would have been very real and *new* in her direct perceptual experience of her  $W[i]$  – are precisely the sort of contents that a person can have direct perceptual experience of in their  $W[i]$  but which they cannot find, identify or hold – through any amount of language learning and use – to be contents of  $W[z]$ .

In the thought experiment, all of the knowledge Mary has of what it is to see colours ahead of being let out of her room – and this knowledge is exhaustive – are contents she has found to be in  $W[z]$  that are *about* seeing colours, as distinct from knowledge of directly perceiving colours as contents of her  $W[i]$ . Here, to say that her knowledge at the level of  $W[z]$  is exhaustive is to say that Mary knows – from her reading, talking and overall communication with others – of *all* of the colour-relevant common features that people have collectively placed into  $W[z]$  by communicating with each other to identify them, name them and otherwise describe them. So any new knowledge Mary acquires when she finally directly perceives colours as contents of her  $W[i]$  must be additional, and *different*, to any of the common features that can be found among the contents she and others have in their  $W[z]$ .

Thought experiments such as Mary’s Room constitute a cognitive procedure by which people can come to realise there are contents in what they directly experience as their  $W[i]$  that they cannot hold through any amount of language learning to be in  $W[z]$ . Yet by each implementing such a procedure, and thus each conceptually isolating such contents, people can – as mentioned above – get as far as generically naming those contents as part of agreeing with each other that such contents can really exist, even though that existence can *only* be at the private level of their respective direct perceptual experiences of their own  $W[i]$ s.<sup>28</sup>

---

<sup>28</sup> Mary’s Room is the best known example of this type of cognitive procedure, but for those who think it too contrived, some straightforward and effective alternative examples have been described by Marc Champagne (21).



The name adopted here for such contents is *qualia*.<sup>29</sup>

Having identified qualia as contents of a person's W[i] that are not in W[z] – and thereby demonstrating a difference between the contents of a person's W[i] and the contents of what they hold to be W[z] – a further difference will now be demonstrated by identifying contents of W[z] that differ from any contents a person can find in their W[i].

Compared to qualia such contents are straightforward to identify and can be considered a subset of common features that can be called *imaginary common features* to differentiate them from the subset of common features described so far, which can be called *concrete common features*.

Atoms, small molecules, Higgs bosons and quarks are examples of imaginary common features. For example, no one has ever directly perceived an oxygen molecule, but through use of language and related cognitive means – particularly in the practice of science – people have come collectively to agree that oxygen molecules exist as ubiquitous and enduring features of our shared physical world, where they may be conceived of as 'hidden' by being too small to be directly perceived as contents of anyone's W[i]. Certainly chemists, physicists and engineers of many kinds consider oxygen molecules to be a common feature in what they each hold to be W[z], having all agreed to this based on their ongoing application of an agreed definition of atoms, agreed measuring procedures and agreed actions where, in implementing those actions, they are able *reliably and successfully* to transform their respective W[i]s to achieve *specific agreed* outcomes.

Having identified qualia, and drawn a distinction between concrete common features and imaginary common features, it is now useful more explicitly to describe the relationship between a person's W[i] and what they hold to be W[z]; i.e. between their W[i] and W[z] levels.<sup>30</sup>

To achieve this it is important to recognise that in normal circumstances – unlike Mary's – a developing infant will have direct perceptual experience of contents of its W[i] *before* knowing which of those contents it will be able – through learning and use of language – to find, identify and name as common features, i.e. as contents of W[z]. So pre-language, an infant's W[z] will be empty of contents.<sup>31</sup> For this reason, at this stage all that an infant will be able to experience in directly perceiving the contents of its W[i] are what can be called *pure qualia*.

It is now helpful to step through what happens next using the colour red as an example. Pre-language, infant Jane will directly experience a given pure quale. She will then learn from others – through the language learning process – to find and recognise this quale as being a concrete common feature that is named, identified and described by use of the word 'red' among members of her language using group. Having learned this, she will then be able to use the word 'red' to tell others, and to learn from others, about something that she originally – and at that earlier stage could *only* – experience as a pure quale.

A crucial concept may now be introduced. It is that *in learning* to recognise that a given specific quale, as she directly perceives it, is *also* a common feature named 'red', Jane will reflexively *cojoin* that specific quale to that specific concrete common feature. In this process Jane will recast what has previously been for her a pure quale, into what from thereon will be for her an *attendant quale*.

---

<sup>29</sup> The term 'qualia' has been given a number of meanings in the literature. In this note the term takes a meaning close to that which Ned Block has identified as the "qualitative or sensational content" of a person's experience of the world as against what he calls the "intentional or representational content" of that experience (22). The meaning here also seems close to that of what Michael Tye – in listing various definitions for the term qualia – calls the "intrinsic non-representational properties" of experience as distinct from the "representational contents" of experience (See Section 1(3) in ref. 20(b)).

With this in mind, it can also be proposed that what have been described above as 'common features' (Section 2) – where we learn to identify, describe and collectively manipulate such common features through learning, use and development of language – can be taken to correspond closely to what Block terms "intentional or representational content" and Tye terms "representational contents" of experience. (If so, when Block defends the existence of qualia by pointing to what he describes as "the fallacy of intentionalizing qualia" (22), his defence can be viewed as pointing to the presumption under the TWP that a person's W[i] can be considered to be in identity with the W[z] of their language using group and asserting that this presumption cannot be sustained.)

<sup>30</sup> Noting that the relationship between a person's W[i] level and the W[r] level has already been described in Section 1.

<sup>31</sup> Even though – as a human infant – its neurology[r] and other physiology[r] may render it precociously *predisposed* to host a W[z] and to populate that W[z] with contents.

Figure 2 provides a graphic similar to a Venn diagram to describe the proposed relationship between a person's physical world at the level of their  $W[i]$  and at the level of  $W[z]$ .

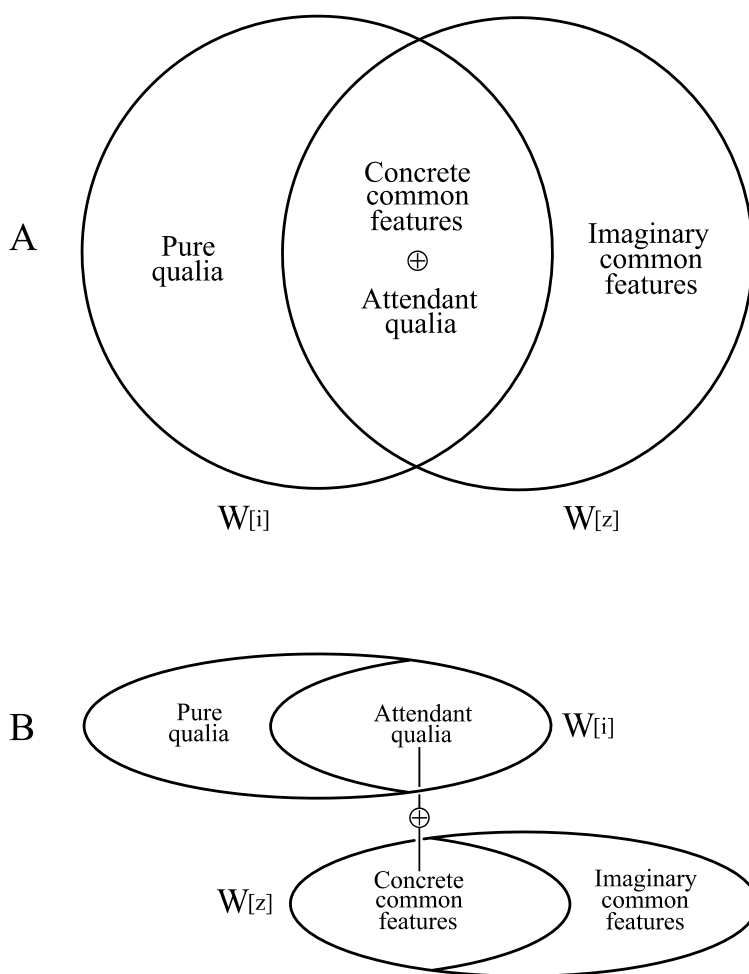


Figure 2

Figure 2 shows a person's  $W[i]$  as a set on the left 'meta-intersecting' with a set on the right of what they hold to be  $W[z]$ . Under definitions provided in the text, contents of  $W[i]$  that are not in the meta-intersection are pure qualia, while contents of  $W[z]$  that are not in the meta-intersection are imaginary common features. Contents of  $W[i]$  and  $W[z]$  that are in the meta-intersection are attendant qualia that have been cojoined – through *language learning operations*,  $\oplus$  – to concrete common features. The lower part of the figure, at **B**, is a partially rotated version of the upper part of the figure, at **A**. This is to illustrate that it is possible, in the abstract, to conceive of the contents of a person's  $W[i]$  as exclusively made up of qualia and the contents of what they hold to be  $W[z]$  as exclusively made up of common features. However, it is crucial to grasp that as part of applying language learning operations, whereby some specific common feature a person can find in  $W[z]$  is identified as – and thereby *conjoined* to – some specific quale they can find in their  $W[i]$ , the person will reflexively 'place' such cojoined concrete common features along with their respective attendant qualia into *both* their  $W[i]$  and  $W[z]$  and will, from then on, *inexorably* experience them as contents of both (see the examples below involving Bill and you).

To get a better feel for what is being presented in Figure 2, it is helpful to return to Mary's Room. What Mary learns can now be cast as follows. At the start of the experiment *all* she knows about colours and about seeing colours, including their respective names, is in the form of imaginary common features she holds to be in  $W[z]$ . But once she is released and directly perceives colours as specific qualia in her  $W[i]$ , she can then apply language learning operations to cojoin one of her newly discovered quales to what had up until then been for her an imaginary common feature named 'red'. With this step 'red' becomes for Mary the name of a concrete common feature.

This mirrors what happens for Jane when, under circumstances more usual than Mary's, she applies language learning operations to cojoin a pure quale she already knows from direct experience of her  $W[i]$  to a common feature named 'red' that she has newly discovered through communication with others. In this case, with this step a pure quale becomes for Jane an attendant quale, where that quale now attends a concrete common feature named 'red'.

Recasting pure qualia as attendant qualia carries inexorable subjective consequences. This can be demonstrated by a thought experiment that considers what a test subject, Bill, experiences when listening to what for him is a completely foreign language. Bill experiences the sounds he perceives as he sits in the audience of a play listening to actors speak this foreign language as pure qualia. Yet all around him native users of the same language will be unable to experience those sounds as pure qualia, only as attendant qualia. For them it will be impossible to experience what the actors are saying *only* at the level of perceived sounds (i.e. only at the W[i] level) because they will find themselves utterly unable – should they ever try to do so – to ‘turn off’ *also* experiencing those sounds as communications about common features (i.e. at the W[z] level).

Similarly, in learning to read you will have lost the ability to experience a word such as ‘vomit’ *purely* as a perception of a pattern of ink on a page, i.e. as a pure quale. Having learned to read, you now *must also* experience that pattern of ink as the concrete common feature named by ‘vomit’. You will always – both before and after learning to read – perceive the same pattern of ink. It is just that after learning to read you inexorably must perceive and experience that pattern as an attendant quale cojoined to a concrete common feature, whereas before you learned to read you could perceive that pattern only as a pure quale, absent any experience of a common feature cojoined to it.

These last two examples involve spoken and written names or descriptions of common features. To correctly grasp the concept of cojointness it is important to recognise that the common features that make up the contents of W[z] are not equivalent to their names or descriptions, but that agreement on a name or description among a group of language users is an agreement that a given name or description *points to* a given common feature to be held in each of those language users’ W[z]. In this respect, for example, either one of the spoken or written names ‘vomit’ points to the one common feature, as does the description – spoken or written – “stuff that people throw up”.

At the same time, for any given language user in that group, they will *also* – in having learned to name or describe a given concrete common feature – have set some specific quale or qualia that they have already experienced as a direct perception to *point to* the name or description that has been adopted for that concrete common feature by members of their language using group.

So when a person such as Jane identifies a pure quale and – in learning a language – sets it to *point to* a specific name or description used by people in that language using group, that act of learning recasts the quale as an attendant quale and cojoins it to the specific concrete common feature which that name or description points to for language users in that group.<sup>32</sup>

In this respect, the names and descriptions used in a language can be understood to form a kind of ‘hinge’. For example, for *all* English language users the word ‘red’ will point to what will – for all of their practical purposes of communication – be the same, specific set of common features<sup>33</sup> in what they all hold to be W[z]. But for any English language user taken as an individual, the word ‘red’ will *also*, by the agency of that user’s *individual* language learning process, have been set to point to – *and* to be pointed to by – members of some specific set of qualia which that user – and that user *alone* – can perceive directly within their own individual and private W[i], and that will therefore remain *unique* to that user.<sup>34</sup>

Once such pointers are set for a person – through their learning and use of language – that person will come to apply them *involuntarily and reflexively* to all of the attendant qualia they directly experience in their W[i]. This is to say that, in the same way that you or I cannot experience direct perception of the attendant quale which is the ink pattern ‘vomit’ without also experiencing the common feature it points to, we will also be unable directly to perceive, say, *actual* vomit as contents of our W[i] without *also* experiencing the common feature which that attendant quale points to.

---

<sup>32</sup> For this reason, an attendant quale could also be called a *pointing quale*, as opposed to a pure quale, which points to no common feature.

<sup>33</sup> The set of concrete common features whose members are shades of red.

<sup>34</sup> This is simply derivative of a fact we each know – already cited in Section 2 – which is that we cannot *directly* subjectively experience any other person’s phenomenal world as they directly experience it.

In other words, to be as clear as possible, when you experience a direct perception of vomit, you will not only experience this as a direct perception of your own, unique, attendant quale for vomit, you will *also reflexively and inexorably* experience it as the concrete common feature English language users call ‘vomit’, without need of any *explicit* presence of that name or description. This is because for a mature language user *all* of the relevant pointing referred to above will have become so reflexive that it will go ‘straight through’ the hinge<sup>35</sup> – i.e. the relevant words or descriptions – to give a *unified experience* of the attendant quale *cojoint* with its concrete common feature.

If this involuntarily unified experience of attendant qualia and their conjoint common features is then understood to be constantly and inexorably applied across *all* of the attendant qualia a person directly perceives as contents of their W[i], then what they will reflexively experience will be unification of their overall experience of their W[i] with an experience of the W[z] of their language using group. For a normally competent language user this reflexively and involuntarily unified experience of W[i] and W[z] will invoke within them a powerful – practically irresistible – cognitive predisposition to experience their W[i] and W[z] as a single, unified physical world shared with all other people: That is, it will invoke in all competent language users a virtually irresistible predisposition tacitly to assume that their W[i] and W[z] are in identity with each other, consistent with their tacit implementation of the TWP, as described in Section 3.

Yet as shown – and as described in caption to Fig. 2 – a person’s W[i] can be conceived of as being at the private level of how they directly perceptually experience pure and attendant qualia and, in this respect, as being quite different to W[z], which can be conceived of as being at the shared level of how that person, as part of a language using group, experiences – and participates with others in manipulating – concrete common features and imaginary common features such as bricks and atoms.

(iii) *Differences between W[z] and W[r]*

A perspective on the differences and relationship between W[z] and W[r] can be gained by considering the history of emergence of the ‘physical worlds’ under discussion, i.e. the history of emergence of the W[r] (the noumenal), W[i] (the phenomenal), and W[z] (the collective) levels.

Under the definitions and assumptions made in Section 1, the first to come into existence will have been the physical world at the level of W[r]. And certainly most people assume that such a world came into existence long before life on earth emerged. This assumption has become so entrenched that it is built into our science-based world view, which holds that a physical world existed while – and long before – the earth itself formed, eons before any human observer could have been present to affect that world or to experience it directly as phenomena in their W[i].<sup>36</sup>

At that time, before the emergence of life, there would have been no terrestrial human or animal information processing arrangement[r]s – specifically, D[r]s implementing RBCs – to give rise to any direct perceptual experiences of W[i]s. But as life on earth emerged and evolved such information processing arrangement[r]s (pIPAs)<sup>37</sup> must also eventually have begun to emerge, with some forming a starting place for further evolution into the pIPAs that we each have within our own B[r]s, and which currently deliver the human experience of phenomenal worlds (W[i]s) under discussion in this note.

It seems likely that neuronal architecture[r]s constituting pIPAs would have begun to emerge in animals long before the emergence of humans, but that it has only been with the emergence of modern humans, and perhaps our closest ancestors, that profound development and use of language – call this *true language* – then also emerged.

If pIPAs emerged in some animals long before emergence of true language, then animals with those earliest types of pIPAs would have experienced some kinds of phenomena; i.e. they would have had some kind of direct perceptual experience of the contents of a W[i]. But for such an animal – and for all animals having pIPAs but no capacity to develop any intraspecies communication akin to true language – there would be no ability to find common features and to develop a W[z].

<sup>35</sup> In this sense the ‘hinge’ will be transparent to mature language users.

<sup>36</sup> The contents of what we agree to have been in the physical world in prehistoric times include the contents of the W[z] of scientists of prehistory and cosmology. The fact that W[z] can have such contents – such as living dinosaurs (a further example of what have been defined above as imaginary common features) – again demonstrates a difference between any given person’s W[i] and W[z], where imaginary common features of the ancient world can never have been directly experienced as phenomena – i.e. experienced at the level of their W[i] – by any person – and can never become concrete common features in any human W[z] (absent a time machine).

<sup>37</sup> Here a pIPA is defined as an information processing system[r] – i.e. a D[r] implementing an RBC, as proposed in Section 1 – that operates on sensory input from W[r] to generate a W[i], where the contents of that W[i] will be directly perceptually experienced as phenomenal experience by some specific physical system[r] – a person[r], animal[r] or perhaps a future robot[r] – that is housing that pIPA as it operates.

Such an animal could be expected to have direct perceptual experiences of phenomenal contents of a  $W[i]$ , but could only ever experience those contents as pure qualia. The creature would have no way of knowing whether any of its fellow species members were themselves having any kind of phenomenal experience or that anything other than their own phenomenal experience might be real.<sup>38</sup>

Accordingly, in the history of emergence of the proposed levels of ‘physical worlds’  $W[r]$ ,  $W[i]$  and  $W[z]$ , the  $W[z]$  level will have been the most recent to emerge – possibly only in humans and our nearest ancestors – and will have required prior emergence of both the  $W[r]$  and  $W[i]$  levels.

The earliest versions of  $W[z]$  to emerge, as our ancestors began evolving towards use of true language, may have had diminished – and less well-articulated – contents compared to our contemporary human  $W[z]$ . In particular, it seems likely that the earliest versions of  $W[z]$  would have contained only concrete common features (CCFs). That is because it seems unlikely that our ancestors – at the time they *first began* to build and use language – would have been able to make and use names and descriptions for things that none of them could directly perceive in their respective  $W[i]$ s.

Regardless, somewhere on the evolutionary path to present day humans an ability to create (imagine), name, describe and talk about imaginary common features (ICFs) must have begun to emerge. To better describe how this might have proceeded, it is helpful to sort ICFs into two kinds. For now call these *unanchored* ICFs and *anchored* ICFs. Of these, it seems likely the first ICFs to emerge would have been unanchored ICFs.

The proposed distinction between an unanchored ICF and an anchored ICF can be drawn by pointing to examples among ICFs that we currently name or describe.

The god Zeus, centaurs, perpetual motion machines and life on other planets are – at time of writing – unanchored ICFs. We can talk about these ICFs, but we are unable to share with each other anything that we can *agree* constitutes *definitive evidence* that such things actually exist, or can have existed, in what we collectively hold to be physical reality; i.e. in what we collectively hold to be the physical world. These examples also show that unanchored ICFs can be further sorted into *hypothetical* ICFs (e.g. life on other planets) or *fictional* ICFs (e.g. centaurs).

Atoms, small molecules and Higgs bosons are anchored ICFs. In these cases people are able to share with each other what they can agree constitutes ‘definitive evidence’ that such things exist, or have existed, in what we collectively hold to be the physical world.<sup>39</sup> In this respect, anchored ICFs can be considered ‘*proven*’ ICFs.<sup>40</sup>

But how do people reach agreement on what constitutes ‘definitive evidence’? For now, simply say that a group of language users can reach agreement on – i.e. can collectively decide – what constitutes definitive evidence that an ICF is physically real by conceiving (imagining), communicating about, and agreeing upon a *rule set* which, when they each apply it in relation to some given ICF, will allow them each to decide whether to hold that ICF to be physically real; i.e. to decide whether or not it actually exists in what they hold to be the physical world. Such rule sets generally entail descriptions of predictive physical procedures.<sup>41</sup>

With this say that agreement can also be reached on ‘higher order’ rule sets – i.e. rule sets about rule sets – conceived of, articulated, and applied by a group of language users in seeking to agree on whether a *hypothetical* rule set is *sound* – i.e. can be agreed on – or *unsound* – i.e. should not be agreed on – as either:

- a means of deciding whether they can consider an ICF to be a proven ICF or not; or
- a means of deciding whether some further hypothetical rule set is sound.

Rule sets can be considered akin to ICFs insofar as an ICF can be agreed to be either hypothetical, proven or fictional and a rule set can be agreed to be either hypothetical, sound or unsound.

---

<sup>38</sup> Some species do have communications among individual members and some such species may have pIPAs more like those of humans than were the first pIPAs to emerge. So for some of those species – for example great apes, and perhaps some other species that operate in groups or packs – a limited kind of  $W[z]$  might exist. But it seems likely, in the absence of true language, that such  $W[z]$ s would be impoverished compared to the  $W[z]$  humans sustain and grow through our ongoing use and extension of true language among our species.

<sup>39</sup> Even though – as for all ICFs – no one can find any qualia within their  $W[i]$  to cojoin to such ICFs.

<sup>40</sup> NB: Here the word ‘proven’ is assigned a specific relatively narrow meaning in its use as part of the term ‘proven ICF’. It is not being ascribed any other, or wider, meaning.

<sup>41</sup> That is, they generally – and virtually always in the sound practice of the physical sciences – entail one or more physical procedures where there is agreement that the physical outcome of those procedures, as directly perceived in each deciding person’s  $W[i]$ , will provide definitive evidence as to whether or not a hypothetical ICF should be considered a proven ICF, a fictitious ICF or even, in some cases, a CCF. (In the latter case, for example, if a life form native to another planet is ever eventually directly perceived by visiting astronauts it will become a CCF. Notably here the agreed rule set applied for reaching agreement that a hypothetical ICF is a CCF entails only the simple physical procedure that members of the language using group can get into a situation (get to the planet) where the hypothetical ICF can (or will fail to) be directly perceived in the contents of their respective  $W[i]$ s.)

Call communication of a rule set an *articulation*, where clearly it will only be possible for members of a language using group to agree on whether a hypothetical rule set is sound or unsound if they literally articulate it in communicating with each other. Articulation of a rule set will almost always take the form of a predictive *explanation* of some set of physical occurrences that reproducibly manifest in the W[i]s of members of a language using group.<sup>42</sup> Whether a rule set is hypothetical, or has been agreed to be sound or unsound will then depend on whether the predictive explanation it offers for such physical occurrences is considered hypothetical, or has been agreed to be sound or unsound.<sup>43,44</sup>

The overall history of emergence of W[r], W[i]s and W[z]s, as proposed, can now be rounded out by further proposing that as W[z]s emerged and evolved among language using groups:

- a) the first ICFs would have emerged – i.e. would have begun to be created and referred to by language using groups – where all such ICFs would have each, at least initially, been hypothetical;
- b) the first rule sets/explanations would have coemerged with the first ICFs, with language using groups conceiving, agreeing upon and then adopting those rule sets/explanations as means of deciding whether a hypothetical ICF should be held to be proven or fictional (or to be a CCF<sup>45</sup>);
- c) where, with this, an ongoing process would have begun entailing the creation of hypothetical ICFs and – where applicable rule sets/explanations could be agreed for those ICFs – sorting them into proven ICFs or fictional ICFs, leading to accumulation in the contents of an evolving W[z] of ever greater numbers of proven ICFs, i.e. ICFs held by members of a language using group to be physically real; and
- d) where this would not have been – and is still not – a linear progression, since new observations might sometimes arise driving emergence of disagreements within a language using group about whether one or more previously agreed rule sets/explanations should – in the light of those new observations – continue to be agreed to be sound (see the example of phlogiston, below).<sup>46</sup>

From here it can be proposed that the history of emergence of W[r], W[i]s and W[z]s has, if only in the last few hundred years, reached a stage where evolution of human W[z]s have led to:

- e) a more-or-less species-wide, single W[z] whose contents include large numbers of ICFs as well as CCFs;
- f) where a great many ICFs have already each – through application of extensive rule sets/explanations long established to be sound – been decided through collective agreement to be either proven ICFs, e.g. atoms; fictional ICFs, e.g. unicorns; or CCFs, e.g. black swans;
- g) where the number of hypothetical ICFs, though still large, is currently almost certainly smaller when compared to the number of proven ICFs we each hold to be in W[z] than would have been the case for people during earlier stages of the evolution of our W[z];
- h) where a key evolutionary process still underway in relation to the contents of our W[z] is a collective effort to seek to agree on which of the ever more abstract hypothetical ICFs we have created, and are still creating, can be determined to be proven ICFs – i.e. can be held to be physically real; and
- i) where, fully interdependent with (h), a process remains underway to collectively seek to agree on which of the often highly abstract, encompassing and higher order rule sets/explanations we have created, and are still creating, can be determined to be sound for the purposes of deciding whether a given hypothetical ICF should be considered proven or fictional.

<sup>42</sup> ‘Almost always’, because – as happens in the course of scientific discovery – some rule set articulations can be made, empirically tested, used and agreed upon *before* an explanation can be developed or agreed upon for them. For all of the purposes of this note however, a simplifying assumption can be made that all articulations of rule sets do – or will – take the form of explanations, and the term *rule set/explanation* will be adopted from hereon.

<sup>43</sup> For example, if a group of chemists were asked why they agree that atoms are proven ICFs – i.e. are physically real – the *explanation* they offer – or point to in the scientific literature – will be an articulation of rule sets that people in the ‘chemistry language group’ (footnote 47) have agreed are sound in order to then agree that atoms are proven ICFs. In articulating the relevant rule sets, their explanation will be *predictive* in that it will entail descriptions of predictive physical procedures in the form descriptions of chemical experiments and their predicted and observed outcomes.

<sup>44</sup> NB: (1) In this note, under this definition, all *explanations* are explanations that people apply to what we call “the physical world” and are articulations, within a given language using group, of rule sets where those rule sets can be hypothetical, or be agreed by that group to be sound or unsound. In this respect an explanation, so defined, can be considered – especially in relation to scientific explanations – to be the same as a hypothetical, sound or unsound *theory*.

(2) Pure prose explanations of a rule set, though they may be adequate for all practical uses of that rule set, may contain imprecisions in their articulation of that rule set *when compared* to rule set articulations in prose plus mathematics. In such cases – where a rule set contains rules about relationships among quantities that can be measured using agreed procedures – rule set articulations using mathematics can be exact, and therefore more predictively precise, than can pure prose rule set articulations.

<sup>45</sup> Where the rule set/explanation for deciding whether a hypothetical ICF could be agreed to be a CCF would be as per footnote 41.

<sup>46</sup> In this proposed process of emergence and evolution of a W[z] within a language using group there will clearly be a deep interdependence between agreement on whether an ICF is proven or fictional, and agreement on whether the rule sets/explanations to be applied in deciding whether an ICF is proven or fictional are themselves sound or unsound. (Hence use of the term ‘co-emerged’ in 5(iii)(b), and see 5(iii)(h) and 5(iii)(i)).

Some examples illustrating this proposed evolutionary history of W[z] are:

A. *The Higgs boson*

There was no ICF called the Higgs boson until circa 1964, when it was created as a hypothetical ICF by Peter Higgs and others. The Higgs boson remained a hypothetical ICF until 2012 when, through the application of rule sets/explanations agreed by people in the ‘physics language group’,<sup>47</sup> agreement was reached that it be accepted as a proven ICF, i.e. that it be held to be physically real.

B. *Phlogiston*

There was no ICF called phlogiston until the early 18<sup>th</sup> century,<sup>48</sup> when it was created as a hypothetical ICF by Georg Stahl and others. For many years, through the development and application of rule sets/explanations agreed by people in the ‘chemistry language group’,<sup>47</sup> agreement was maintained among members of that group that phlogiston was a proven ICF; i.e. it was actually held by chemists to be physically real. However, as rule sets/explanations agreed by chemists evolved – consistent with (d) above – agreement that phlogiston was a proven ICF progressively eroded until eventually, some hundred years after its creation, it came to be agreed that phlogiston was a fictional ICF.

C. *God*

There is disagreement over whether the hypothetical ICF ‘a single god of all things’ (God) should be held to be a proven ICF – i.e. to be physically real – or a fictional ICF.<sup>49</sup> This includes disagreement over which rule sets/explanations should be applied to yield what could be agreed to be definitive evidence that God is physically real. Rule sets/explanations entailing no well defined predictive physical procedures, and that are reliant on concepts such as faith, are particularly contentious.<sup>50</sup>

The history proposed above of serial emergence of W[r], then W[i]s, then W[z]s, followed by progressive convergence of those W[z]s onto a single, species wide human W[z], provides a basis for exploration of the differences, and the relationship, between W[r] and W[z].

In particular, the trajectory described by this history allows extrapolation of the evolution of W[z] to a potential endpoint. Through consideration of (a) through (i) above, this endpoint can be proposed – *if only as an ideal* – to be:

- j) a single, species-wide W[z] whose contents are universally agreed to be either proven ICFs, fictional ICFs or CCFs, and with the number of hypothetical ICFs approaching zero;
- k) where this agreement on contents is sustained through agreement on a single, well defined, universally applicable rule set/explanation comprised of logically consistent,<sup>51</sup> mutually reinforcing higher and lower order rule sets/explanations; and
- l) where, into the future, no new observations are encountered that raise doubts about the universal agreement referred to at (k).

---

<sup>47</sup> The ‘physics language group’ is simply defined as the language using group people participate in when they communicate and reach agreements in the area of physics. The ‘chemistry language group’, a wider ‘science language group’ and other such groups can be defined similarly.

<sup>48</sup> Although ideas leading to the proposal of a substance named phlogiston were introduced by Johann Becher in the mid 17<sup>th</sup> century.

<sup>49</sup> Or, with respect to the very rare cases where people believe they have directly witnessed God, a CCF.

<sup>50</sup> It is clear that before the rise of modern science, faith-based rule sets/explanations, including those reliant on religious or spiritual beliefs, were central to what then constituted communicable human understanding – i.e. explanations – of the contents of the physical world (i.e. the contents of a W[z]). At that stage in the evolution of any W[z], ICFs such as gods and spirits were generally considered within various language using groups to be the creators/causes of all of the physical objects and events that any person could perceive directly as contents of their W[i] and also communicate about as CCFs and their movements at the level of the agreed contents of a W[z]. But disagreement about such ICFs, and the rule sets/explanations underpinning these, would have arisen when different language using groups began to mix. How to decide which gods were real and which religious rule sets/explanations were sound? Most likely, as new language using groups emerged from merger of older groups, new – or at least modified – religious rule sets/explanations (faiths) and ICFs (gods, spirits or other supernatural entities) emerged to allow merger of W[z]s. Moreover, it is empirically clear that – as an evolutionary process and *on balance* – such syncretic modifications and mergers have amounted to an ongoing rationalisation of religious and spiritual beliefs leading to the invention and availability of religions seeking agreement to the existence of a single ICF (god) and to a rule set/explanation (faith) supporting belief in that single god (23). Importantly, the invention of a single ‘unified’ god and an attendant rule set/explanation providing for that god to be the one creator/cause of all physical things and processes would seem to make that god an excellent midwife to the birth of modern science. That is because modern science also seems to claim (and essentially to take it on faith) that the physical world will indeed prove to be a single unified system operating under, and subject to, a single unified rule set/explanation (as per main text (j) – (l) above). Here the evolution of W[z]s – if roughly construed as entailing progression from animisms to polytheisms to monotheisms – can be seen as a process whereby spiritual creators/causes of physical objects and events have – through successive versions of W[z]s – been incrementally rationalised down to a unity. Importantly, this seems to have set the stage for a further step – already taken by many scientists and others – which is to agree that the ultimate ICFs and rule sets/explanations required fully to describe the single, unified the physical world that such people hold we live in, will eventually stand seamless and complete, without any need of – or perhaps any place for – a god or any other spiritual ICFs or their attendant faith-based rule sets.

<sup>51</sup> Where what constitutes logic can itself be considered a higher order rule set.

This proposal is not as contentious or abstract as it may seem. At this stage in history it is not difficult to conceive of such an endpoint, and there seems no reason why it cannot be achieved.

This is because the emergence and evolution of science<sup>52</sup> as it applies to what we hold to be the physical world clearly suggests that members of the science language group,<sup>47</sup> through their collective practice of experiment and theory making, may well be converging on successfully creating and agreeing to a single, universally applicable rule set/explanation – a successful ‘theory of everything’ – as described at (k).

If the sum of currently agreed rule sets/explanations – which can loosely be called ‘the corpus of scientific knowledge and reasoning about the physical world’ – can indeed be developed and agreed upon to that extent, then in that ideal case its application should *always* allow agreement to be reached among the science language group on whether any arising hypothetical ICF should be considered proven or fictional.

And if, in turn, such a ‘scientific world view’ – based on a *completed* ‘corpus of scientific knowledge and reasoning about the physical world’ – were to be universally accepted, this would deliver a single, species-wide  $W[z]$  as described at (j). In this ideal case, it could then validly be agreed that the science of physics, including cosmology, had reached completion<sup>53</sup> and that an ideal form of relationship between  $W[z]$  and  $W[r]$  would apply.

Yet even in this ideal case,<sup>54</sup> the relationship between  $W[r]$  and  $W[z]$  could not be one of *identity*, as tacitly assumed under the TWP. People in the science language group may agree, and may come genuinely to believe, that ICFs such as atoms and many rule sets/explanations – including physical laws – actually exist independently of whether human beings exist (i.e. exist in, and at the level of,  $W[r]$ ), but strictly speaking such beliefs simply reflect a wholesale adoption and implementation of the TWP. Certainly, in order for the TWP to *work* – and it would work perfectly for people in the ideal case under discussion – the CCFs, proven ICFs, and rule sets/explanations of an ideal  $W[z]$  would have perfectly to *correspond* to what exists in  $W[r]$ <sup>55</sup> – and might *in that special sense* be said to ‘exist’ – but fundamentally, they could only ever be our *descriptions* of what exists in  $W[r]$ , they could never actually *be* what exists in  $W[r]$ .

A shorthand way of supporting this assertion is to imagine that everyone, including all scientists, are wiped out by a virus. Under the assumptions adopted at the outset of this note,  $W[r]$  would carry on *oblivious*, unimpacted except for an absence of human physical agency.<sup>56</sup> But there would be no human  $W[i]$ s and there could be no human  $W[z]$ , and with this there could be no CCFs, ICFs or rule sets/explanations.<sup>57</sup>

## 6. Supervenience

The relationships between  $W[r]$ ,  $W[i]$  and  $W[z]$  can be viewed from a structural and information processing perspective. The proposed structural relationship among the three is shown overleaf in Figure 3.<sup>58</sup>

Now consider the further proposals, already broadly covered above, that:

1.  $W[r]$  is as described in Section 1, including with reference to the structure shown in Fig. 1. It is where people’s physical body[r]s – their  $B[r]$ s – exist and house their brain[r]s and where, within those brain[r]s, people’s respective  $D[r]$ s process incoming sensory information through application of their own individual RBCs.

Noting that with this it is being proposed that information can be, and is being, processed in  $W[r]$ :

- (a) where for any system[r], made up of thing[r]s, such information processing will take the form of changes in state over time of that system[r], where such changes are defined as changes in relations among and/or within those thing[r]s.

<sup>52</sup> In particular the physical sciences, including particle physics, gravitational physics and cosmology.

<sup>53</sup> Completion in the larger sense of being able to fully explain and, to the maximum extent possible predict, the contents and behaviour of the physical world. This form of completion would encompass the type of completeness described, for example, by David Papineau (24).

<sup>54</sup> And notwithstanding that it remains an open question whether an ideal  $W[z]$  – as described at (j), (k) and (l) above – might ever actually be achieved.

<sup>55</sup> The specifics of this *correspondence* requirement for the TWP to work are described in Section 4, and in particular Section 4 at (iii).

<sup>56</sup> Where here human physical agency refers only to the physical outputs of living human  $B[r]$ s into wider  $W[r]$ , including all motor output, as shown in Fig. 1.

<sup>57</sup> True, in this scenario libraries might remain filled with articulations of the human knowledge of CCFs, ICFs and rule sets/explanations required to allow an able group of readers to generate an effective  $W[z]$ , but with no one left to read and understand the communications in those books, no such  $W[z]$  could be generated.

<sup>58</sup> Figure 3 simply extends the structural relationship between a  $W[i]$  and a  $W[z]$  described in Fig.2 to encompass the structural relationship between a  $W[i]$  and  $W[r]$  described in Section 1.



2. A W[i] is generated within a person's brain[r] through the processing of sensory information[r] by their D[r] as described at (1) immediately above. The W[i] being directly experienced by a person as their phenomenal world will *dynamically map* the state of their B[r] in W[r].

No information can be processed *within* a person's W[i] – i.e. at the W[i] level – since a W[i] is purely phenomenal.<sup>59</sup> But changes in state in some system[r]s in W[r] will be *mapped* by operation of a person's D[r] and RBC to give a directly perceived evolution of phenomena in their W[i].

A key implication of this – and of the more detailed description given in Section 1 of how a person's W[i] can be generated out of the information processing operations of their D[r] – is that everything in a person's W[i] – i.e. all that they directly perceive as contents of their phenomenal world – will supervene on contents of W[r], including in particular, on information held in their evolving brain[r] structures and brain state[r]s.

3. A W[z] is founded on, and depends upon:

- (a) development in people of their own respective W[i]s, in concert with;
- (b) people's native human linguistic ability to refer to the contents of their respective W[i]s to learn and to use – including through the process of cojoining qualia to common features – a pre-existing language already in use by members of a language using group,<sup>60</sup>

such that, by this means, people are able to join in the use of, propagation and evolution of that language, and therefore of the W[z], shared by that group.

It is proposed here that all information processing associated with learning and using language takes place solely within W[r] – i.e. at the W[r] level – by information processing system[r]s operating in conjunction with, but extending well beyond, those system[r]s in a person's brain[r] that generate their W[i]. These extended system[r]s – including neurological system[r]s corresponding in W[r] to what the neuroscience of our W[z] describes as “language areas” – seem sure to reside mainly within people's individual brain[r]s.<sup>61</sup>

Beyond this, and crucial in the scheme proposed, people's brain[r]s communicate with and inform each other as they learn, use, modify and extend language.

This means that the information and information processing system[r]s that constitute and deliver to users the contents of the W[z] of their language using group, will exist in W[r] as a more-or-less single,<sup>62</sup> dynamic, information storage and processing system[r] that is networked and resides across the brain[r]s of members of that group, with information passing backwards and forwards between the brain[r]s of people in the group as speech sound[r]s, bodily gesture[r]s, text[r] and so forth, *all* at the W[r] level.

Consistent with this, it seems clear there is no way for information to be processed *within* a language using group's W[z] – i.e. at the W[z] level – since, as defined, the W[z] level is made only of agreements or proposed agreements on names and/or descriptions of CCFs, ICFs and of agreements or proposed agreements on rule sets/explanations.<sup>63</sup> Hence, as with the W[i] level, the W[z] level will lack any substrate within which information can be processed.

If so, then everything that happens in and through language – including all interpersonal communications and the contents of any W[z] – will supervene on information and information processing that is held in, and takes place within, system[r]s in W[r].

Together with the supervenience of any W[i] on W[r] – as per 6(1) above – this means that in the scheme proposed in this note, *everything* that takes place within any W[i] or any W[z] will supervene on *something* that takes place in W[r]. In other words *all* content and its evolution at the W[i] level, and *all* content and its evolution at the W[z] level, will supervene on some or other content and its evolution at the W[r] level. In this respect the scheme proposed here can be considered a form of physicalism.<sup>64</sup>

<sup>59</sup> This point was made in Section 1, and specifically in footnote 11, and is held to be self-evident.

<sup>60</sup> The assertions 3(a) and 3(b) flow naturally from consideration of the description and definition of W[z] provided in Section 2 and elaborated in Section 5(ii), and through consideration of the evolutionary history of W[z]s proposed in Section 5(iii).

<sup>61</sup> ‘Mainly’, because these extended information processing system[r]s have very recently begun to encompass computer[r]s and other such thing[r]s in W[r] that sit outside people's brain[r]s and B[r]s.

<sup>62</sup> ‘More-or-less’, because a person can be cut off from communications with some or all others in their language using group for long or short periods of time, and also because no individual member of any language using group will have full mastery of their language or full knowledge of all CCFs, ICFs or rule sets/explanations that are current in that language and its associated W[z].

<sup>63</sup> Even though it seems equally clear that some of the rule sets/explanations created, articulated and agreed upon by language using groups – especially those expressed using mathematics and subject to quantitative verification using agreed measuring procedures – are able most effectively to *describe* how information is processed in W[r], and can sometimes do so with great precision.

<sup>64</sup> At least to the extent of being a *supervenience physicalism* (25), where the contents of the W[r] level and their evolution are considered to be the ground – i.e. the objective reality – upon which, for any person, all contents and their evolution within their W[i] and W[z] levels will depend. (Noting here that if a person wants to *call* the contents of their own W[i] and its evolution ‘real’ – i.e. call ‘real’ the state and evolution of the pure and attendant qualia that they directly and with such great immediacy perceive to exist within their own private W[i] – this can be considered reasonable, *provided it is understood* that when such a person says something is ‘real’ in this context what they are *actually* saying is that it is something that is ‘real to me as a subjective experience’.)

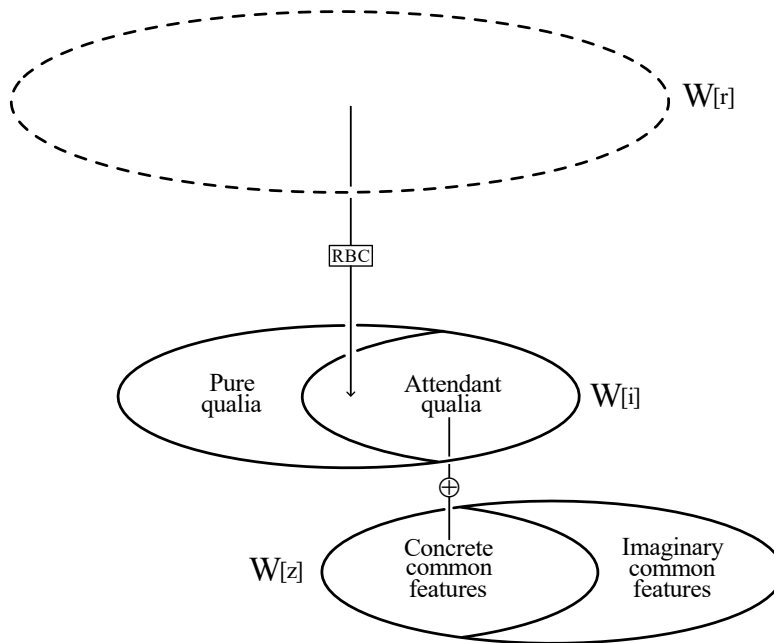


Figure 3

Figure 3 describes the proposed structural relationship between  $W[r]$ ,  $W[i]$  and  $W[z]$  for a person. It shows some of the contents of  $W[r]$ <sup>65</sup> being ‘mapped’ into the contents of a person’s  $W[i]$  through operation of their  $D[r]$  as it applies an RBC to incoming sensory input $[r]$ .<sup>66</sup> These latter contents are shown as qualia, which are the contents of a person’s direct and private perceptual experience of their phenomenal world, which is their  $W[i]$ . These qualia are cast either as pure qualia or attendant qualia where – as described in Fig.2 and Section 5(ii) – for any given person attendant qualia will be cojoined to concrete common features to form a meta-intersection between the person’s  $W[i]$  and what they hold to be  $W[z]$ . What they hold to be  $W[z]$  is shown as containing common features, where these are cast either as imaginary common features or as concrete common features and where – as described in Fig.2 and Section 5(ii) – for any given person concrete common features will be cojoined to attendant qualia to form a meta-intersection between what that person holds to be  $W[z]$  and to be their  $W[i]$ .

## 7. Why bother to distinguish between $W[r]$ , $W[i]$ and $W[z]$ ?

Sections 3 and 4 proposed that with the emergence of language people began to apply the three world praxis (TWP) and that we evolved into doing so because it gives our species an immense survival advantage.

Recall that Section 3 describes the TWP as follows:

For everyday purposes we each operate on a compound and *tacit* assumption – essentially, a praxis – that  $W[r]$ ,  $W[i]$  and  $W[z]$  are identical to, and can be held to be in identity with, each other. That is, we operate on a tacit working assumption that  $W[r]$ ,  $W[i]$  and  $W[z]$  can be held to be the same thing.

In maintaining this praxis we each, for all practical purposes, tacitly assume – and act – as if:

- α) the physical world as we each experience it phenomenally, in the form of our  $W[i]$ , is at any given moment a direct veridical experience of – and may be held to be a world identical to, and in identity with – part of a single unified physical world that exists independently of all of us, at the noumenal level, i.e.  $W[r]$ ; and that
- β) the physical world as we each experience it phenomenally, in the form of our  $W[i]$ , presents itself in exactly the same way phenomenally – and may be held to be a world identical to, and in identity with – the physical world as it is experienced phenomenally by all other people in the form of their respective  $W[i]$ s.

Where, to compress (α) and (β) into more familiar terms:<sup>67</sup>

- γ) We each assume that every person’s  $W[i]$  – including our own – provides us each as individuals with the same direct, veridical experience of a single, unified physical world that exists independently of all of us.

Yet if, as Sections 3 and 4 claim, the TWP can be, and is being, implemented so fully and successfully why not simply continue to presume that its underlying assumptions are entirely correct and that what have been described as  $W[r]$ ,  $W[i]$  and  $W[z]$  can all be regarded as the same thing?

<sup>65</sup> Where the dashed line is to indicate that the set of contents of  $W[r]$  may be boundless and where, at the least, the overall contents of  $W[r]$  can be expected to be far greater than the subset of contents of  $W[r]$  that may be mapped at any given time into the contents of any given  $W[i]$ .

<sup>66</sup> In accord with the arrangements described in Section 1.

<sup>67</sup> As per footnote 19.

Certainly to draw no distinction between  $W[r]$ ,  $W[i]$  and  $W[z]$  is consistent with the ‘common sense’, naïve realist way that people normally relate to what we view to be the physical world. And maintaining that  $W[r]$ ,  $W[i]$  and  $W[z]$  should all be considered the same thing is consistent with respected views of many philosophers who consider themselves direct realists.

But apart from the demonstration of means to differentiate between  $W[r]$ ,  $W[i]$  and  $W[z]$  provided earlier in this note, there are strong reasons to use a conceptual framework that explicitly describes what we call the physical world as a composite of conceptually separable  $W[r]$ ,  $W[i]$  and  $W[z]$  levels.

For example, as demonstrated in Section 5:

- drawing a distinction between  $W[r]$  and  $W[i]$  allows for a relatively clear cut description of the nature of our subjective experience of perceptual illusions, dreams and hallucinations;
- drawing a distinction between  $W[i]$  and  $W[z]$  allows for recognition and a relatively clear cut description of the nature of our subjective experience of qualia, and also of our ability to hold to be physically real things that we know no person has or will ever directly perceive, such as oxygen molecules, quarks or Higgs bosons; and
- drawing a distinction between  $W[r]$  and  $W[z]$  allows for a relatively clear cut description of how our agreements on what we hold to be the contents of the physical world – i.e. our agreements on what we collectively hold to be *physically real*, up to and including what our scientists hold to be physically real – are formed through learning and use of language (including of agreed measuring procedures), and can *change* as we acquire new experience/empirical evidence<sup>68</sup>

where *all* of these descriptions – from describing our subjective experience of physical phenomena through to describing our scientific knowledge of what we hold to be an observer independent physical world – fall here within one *unified*<sup>69</sup> and largely empirically supported conceptual framework.<sup>70</sup>

Further to this, a persuasive reason to use a conceptual framework that explicitly describes the physical world as comprised of conceptually separable  $W[r]$ ,  $W[i]$  and  $W[z]$  levels is that recent progress in neuroscience and philosophy of mind strongly suggests that answers to the question of how subjective states – including overall consciousness – can be generated require that a distinction be drawn between  $W[r]$  and  $W[i]$ .

In particular, descriptions of a person’s direct experience of a phenomenal world (a  $W[i]$ ) in relation to, but *as distinct from*, something very like the noumenal world ( $W[r]$ ) described in this note, seem essential – even if this distinction is understated or left implicit in the literature – to both predictive processing (6, 7) and phenomenal self-model (3, 4, 32) accounts of how a person’s subjective experience of the physical world is generated.

(i) *Predictive processing accounts of phenomenal experience distinguish between  $W[i]$  and  $W[r]$*

A proposal common to all predictive processing accounts of a person’s phenomenal experience is that such experience is generated through establishment, maintenance and operation within that person’s brain of a hierarchical processing system that:

- takes sensory input flowing into it from “hidden states” that arise and evolve in their brain’s “environment” and which constitute the “hidden causes” of that sensory input; and
- applies a *generative model* to that input to maintain evolving internal states in the person’s brain that are subjectively experienced by the person as evolving phenomena, where such phenomena are representations of those hidden states.

---

<sup>68</sup> Witness the rise and fall of phlogiston, as per Section 5(iii)B.

<sup>69</sup> Where by ‘unified’ is meant ‘single and internally logically consistent’.

<sup>70</sup> Our standard worldview will not admit to an encompassing and unified framework, but instead is forced – through our tacit adoption of the TWP’s underlying assumption that  $W[r]$  and  $W[i]$  (and  $W[z]$ ) should all be seen as a single *undifferentiable* physical world – to contain an insurmountable conceptual discontinuity between that which we then call “the physical” and that which we call “the mental”.

This impasse has been well described by Jenann Ismael (7), as follows:

“... the central problem: how to bring the view from within, on which I am the frame of the world, the unrepresented representer who contains the whole of it, together with the view from without, on which the world is the frame, and I am somewhere inside the picture, an undistinguished thing among things. In the first, the world is in me; my consciousness is the canvass on which it is painted, everything from the sun and moon to the most distant stars and the smallest insect. In the second view, I am a tiny, transitory centre of a universe that extends beyond me on every side. Both pictures are internally consistent, but together they seem to present an impossible Escheresque construction, each containing the other, confounding attempts to get a unified, overall vision of reality.”

To quote Karl Friston, a key proponent of the predictive processing paradigm (13):

“The theme underlying the Bayesian brain and predictive coding is that the brain is an inference engine that is trying to optimise probabilistic *representations* of what caused its sensory input.”<sup>71</sup>

where that optimisation is achieved through ongoing refinement of a generative model, which is described as follows in a further paper co-authored by Friston (26):

“A generative model is a probabilistic model ... of how sensory observations are generated. It is a statistical mapping from *hidden causes* which include external states of – or causes in – *the environment* to sensory observations.”<sup>71</sup>

The literature of predictive processing refers repeatedly to such “hidden causes”, “hidden states” (e.g. 26, 27, 28, 29, 30), and to an “environment” (13, 26, 27, 30) in which such hidden states/causes are claimed to exist. In some of the literature this environment is referred to as “out there” (26, 31) or as “the distal world” (10). Note then that the predictive processing paradigm logically requires<sup>72</sup> that the brain doing the relevant information processing must – to receive the sensory input required – be *itself* embedded in this “environment”, and be *itself* among the contents of any proposed “distal world” that is “out there”.

It is straightforward to translate the terms given in quotes above into terms used in this note. The “environment”, “out there” and “the distal world” must mean what is meant in this note by W[r], which will then be where a person’s brain[r] processes information, including through the neurological[r] structures that implement predictive processing.<sup>73</sup> Moreover, the literature of predictive processing can be seen as proposing that predictive processing – together with the neurological[r] structures that sustain it – is the means by which a person’s brain[r] learns and implements an RBC, as described in Section 1 of this note, enabling it to generate that person’s experience of phenomena, and of a phenomenal world, W[i].

The relevant literature also proposes that through such processing – provided “all is going well” – these phenomena will, for all of the *practical purposes* of a person’s interaction with the physical world – i.e. interaction of their brain[r] with its environment in W[r] – constitute highly reliable representations at the W[i] level of the relevant hidden states in W[r]. This seems clearly to be what another key proponent of the predictive processing paradigm, Andy Clark, (11) means when he says:

“... what we perceive is (when all is going well) the structured external world itself. But this is not the world ‘*as it is*’, where that implies the problematic notion ... of a world represented independently of human concerns and human action repertoires. Rather, it is a world parsed according to our organism-specific needs and action repertoire.”<sup>74</sup>

“ We encounter our world in perception ... because brains like ours are statistical engines able to lock on to non-linearly interacting causes whose signatures may sometimes be deeply buried among sensory noise and energetic flux. The result is that the agent-salient structure of *the distal realm* becomes reflected in both the large-scale shape and ... the spontaneous activity patterns of the neural architecture itself.”<sup>75</sup>

Given all this, it is difficult to conceive how proponents of predictive processing could hold that the hierarchical processors they discuss could reside in brains situated anywhere other than within an “environment” that can be equated to what this note describes explicitly as W[r], or that the representations that those processors each generate could manifest as phenomenal experience anywhere other than within a ‘proximal realm’ that can be equated to what this note describes explicitly as a W[i].

It is also fairly clear from the above that the predictive processing paradigm requires – even if its proponents leave this implicit – that the “environment” within which the brain is situated be viewed as distinct – i.e. as conceptually separable – from the realm of phenomenal experience (of “sensory observations”) that Friston describes in his second quote above as *mapping* “external states” that exist in that environment.

---

<sup>71</sup> My italics.

<sup>72</sup> Although this seems never to be clearly stated by its proponents.

<sup>73</sup> Noting that this claim is consistent with the proposal made in Section 6 that the only level within which information can be processed is W[r].

<sup>74</sup> My italics, Clark’s parentheses.

<sup>75</sup> My italics.

(ii) *Phenomenal self-model accounts of phenomenal experience distinguish between W[i] and W[r]*

A proposal common to all phenomenal self-model accounts (e.g. 4, 6, 32) of a person's subjective experience is that what they perceive to be their physical body as it sits within and interacts with what they perceive to be their physical environment – which is what they hold to be the physical world – is a model generated within their brain through its processing of incoming sensory information. To quote Thomas Metzinger, a key proponent of the phenomenal self-model paradigm (5):

“The human brain can be compared to a modern flight simulator in several respects. Like a flight simulator, it constructs and continuously updates an internal model of *external reality* by using a continuous stream of input supplied by the sensory organs... It integrates sensory-input channels into a global model of reality, and it does so in real time. However there is a difference. The global model of reality constructed by our brain is updated at such great speed and with such reliability that we generally do not experience it as a model. For us, *phenomenal reality* is not a simulational space constructed by our brains; in a direct and experientially untranscendable manner, it is the world we live in.”<sup>76</sup>

Clearly that which Metzinger considers to be “external reality” can be equated to what this note describes as W[r] and what he calls “phenomenal reality” can be equated to what this note describes as a W[i]. It is also evident from the quote that the phenomenal self-model approach requires that “external reality” and “phenomenal reality” be conceived of as distinct – i.e. as conceptually separable – from each other. As part of this, and as with the predictive processing paradigm, it is also clear from the quote that the phenomenal self-model paradigm logically requires that the brain doing the relevant information processing must itself be situated within this “external reality” – i.e. it must exist as a brain[r] – since this is the only way that it will be able to receive the sensory input it needs to build and maintain an internal model of that reality.

The proponent of phenomenal self-model paradigm who perhaps comes closest to drawing an explicit distinction between what are described in this note as W[r] and W[i] is Antti Revonsuo, who does so at the level of an explicit distinction between what are called in Section 1 the phenomenal body, B[i], and the noumenal body, B[r], where each of these is clearly what Revonsuo respectively refers to in the following quote as “the phenomenal body image” and the “physical body” (33):

“The lived body, or *the phenomenal body image*, is not identical with the organism's body of flesh and bones... The phenomenal body image is a construction at the phenomenal level of organization in the brain, whereas the physical body of flesh and bones surrounds the brain and is in tight causal contact with it. When awake, the state of the physical body of course modulates the state of the lived body at the level of phenomenal organization, just as the external visual stimulus fields modulate the states of the subjective visual field at the phenomenal level, but in no way are the two ever identical... The thoroughgoing distinction between the physical body and the phenomenal body image (or virtual body) is one of the most difficult facts to take in.”

Revonsuo also describes a “phenomenal world” generated within “the brain” (34) where, although Revonsuo leaves it implicit, it is clear that the brain he describes must be a brain[r] that exists within the “physical body” that he describes in the quote above and which, in turn, must be situated within a wider physical world equivalent to that described here as the noumenal world, W[r].

Notably, Anil Seth has combined the predictive processing and phenomenal self-model paradigms in seeking better to describe ‘consciousness’, and his approach is also open to the line of analysis used in relation to the quotes of others given above.<sup>77</sup>

---

<sup>76</sup> My italics.

<sup>77</sup> Seth writes (35):

“Really try to think about what it is like up there, sealed inside the bony vault of the skull, trying to figure out what's out there in the world. There's no light, no sound, no anything – it's completely dark and utterly silent. When trying to form perceptions, all the brain has to go on is a constant barrage of electrical signals which are only indirectly related to things out there in the world, whatever they may be... How does the brain transform these inherently ambiguous sensory signals into a coherent perceptual world full of objects, people and places? In Part Two of this book, we explore the idea that the brain is a prediction machine, and that what we see, hear and feel is nothing more than the brain's ‘best guess’ of the causes of its sensory inputs. Following the idea all the way through, we will see that the contents of consciousness are a kind of waking dream – a controlled hallucination – that is both more and less than whatever the real world really is.”

This passage can be analysed just as has been done for those already quoted in the text. Clearly when Seth talks here of a brain, it can only be viewed as a brain[r], in a skull[r], where “out there” is W[r]. And when Seth talks of a perceptual world, it can only be viewed as a W[i] “full of” object[i]s and people[i].

Looking past the fact that the predictive processing and phenomenal self-model paradigms seem clearly to require that a conceptual distinction be drawn between a physical world that is “out there” – i.e. a  $W[r]$  – and a phenomenal world through which a person experiences that physical world – i.e. a  $W[i]$  – it is useful to ask head on:

1. Are there any problems with describing/conceptualizing the physical world as existing for each of us at three differentiable levels – i.e. with conceptually unbundling the physical world into  $W[r]$ ,  $W[i]$  and  $W[z]$  – apart from the strong discomfit we may feel at having to suspend our evolutionarily driven faith in the TWP and the assumptions shown in this note to underpin it?

It is hard to think of any. All that is being done here is to explicitly describe how we each – as individual subjects and as members of language using groups – actually come to encounter, learn and communicate about, and to singly and cooperatively interact with, what we each call the physical world. And with this, all that is being asked is an admission that taking the TWP to be literal truth is an oversimplification – at least for the purposes of philosophy of mind and cognitive science<sup>78</sup> – of how we should conceive of what we call the physical world.

Moreover, as discussed earlier,<sup>79</sup> this approach offers a way of describing the physical world that is able to encompass within a single, unified conceptual framework:

- ourselves as perceivers: including our phenomenal experience of dreams, illusions, hallucinations, and of qualia, and all of the phenomena we experience as normal contents of our physical environment as we go about our daily lives;
- ourselves as actors: as we interact with, and act to alter, an observer independent physical world that contains all physical things at the ontological level – including our ‘actual’, as opposed to ‘perceived’, bodies and their brains – and where all information is processed; and
- all of the things that we, and our science, hold to be physically real, including those things that we know no person will ever be able to experience as phenomena.

Beyond this, it can then be asked:

2. Are there further advantages to describing/conceptualizing the physical world as existing for each of us at three differentiable levels – i.e. with conceptually unbundling it into  $W[r]$ ,  $W[i]$  and  $W[z]$  – beyond those already discussed?

There are. One such advantage is that use of the conceptual framework described in this note can provide clarifying insights into the Mind-Body Problem. For example, when applied to questions such as<sup>80</sup>

3. (a) what are mental states and what are physical states;  
(b) is one class a subclass of the other, so that all mental states are physical, or vice versa;  
(c) or are mental states and physical states entirely distinct?

it allows the terms used to be untangled from the tacit assumptions underlying the TWP to provide a clearer picture of what is being asked, along with some answers, as follows.

First, the term ‘physical’ as it is used in questions 3(a)-(c) should, as a corrective step, be taken to refer *only* to the contents of  $W[r]$ , and the word ‘mental’ should, as a corrective step, be taken to refer to:

4. the contents of a person’s  $W[i]$ , which will be all of the contents of their subjective phenomenal experience of a physical world and which – by maintaining the TWP – they will generally take to be, and will call, “the physical world”

*plus*

5. the contents of what can be called  $M$ , where  $M$  can be defined as all of the contents of what a person subjectively experiences that are *not* part of their experience of a  $W[i]$  – so, for example, emotions and thoughts will be contents of the person’s  $M$  – which is what a person will generally take to be, and will call, their “mind” or their “mental world”.

<sup>78</sup> And perhaps for some areas of physics – see Ismael (36, 37) and Mermin (38) – although this possibility is not pursued or claimed in this note. (Regarding Ismael’s proposals at (36) it is nevertheless intriguing to ask whether what she calls Ur-space might be considered to be the ontological dimensionality of  $W[r]$ , and what she calls Our-space considered the familiar phenomenal dimensionality – the three dimensions of space and one of time – we each experience as the dimensionality of our  $W[i]$ , and that we then collectively describe, communicate about, and do science in relation to, as  $W[z]$ .)

<sup>79</sup> See page 19.

<sup>80</sup> These questions are taken from Section 1.1 of <https://plato.stanford.edu/archives/fall2020/entries/dualism/> (39) which also provides a clear description of the Mind-Body Problem.

So in 3(a) above, all of the ‘physical states’ referred to should be considered contents of W[r] and all of the ‘mental states’ referred to should be considered contents either of a person’s W[i] or of their M.

It can then be proposed that the contents of what a person *calls* the ‘physical world’ in referring to their W[i], *and* the contents of what a person *calls* their ‘mental world’ in referring to their M, are *both* generated by information processing systems operating in that person’s brain[r]. The generation of a W[i] in this way has already been described,<sup>81</sup> and it should not be difficult for readers to accept that the contents of what a person calls their ‘mental world’ or ‘mind’ are also generated by information processing systems in their brain[r].

Crucially then, in this scheme the key difference between the contents of a person’s W[i] and the contents of their M is *not* that the contents of their W[i] are ‘physical states’ while the contents of their M are ‘mental states’ – since in this scheme the contents of their W[i] and of their M are *all* mental states – the key difference is that the person’s brain[r] generates a W[i] such that all of its contents are directly experienced by the person as distributed in a three dimensional space, which they call “physical space” – but should, as a corrective step, be called a *phenomenal space* – whereas their brain[r] generates M such that its contents are experienced non-spatially, i.e. the contents of their M are *not* directly experienced as contents distributed in any form of space.

Core facets of the Mind-Body Problem arise because, in maintaining the TWP, people assume their W[i] to be in identity with W[r], the observer-independent ‘physical world’. Thus, they take their W[i] to be physical reality, and with this come to hold that all of the contents of physical reality must exist in phenomenal space.<sup>82</sup>

This is an error that leads some to reason that the contents of their M, which they cannot find anywhere among the contents of phenomenal space, must exist somewhere other than in physical reality – perhaps in their ‘mind’ or in a ‘mental world’ – that defies inclusion in physical reality, and whose connection to physical reality is unclear. This leads them to ‘dualism’.<sup>83</sup>

‘Materialism’ can be viewed as a reaction to dualism by those who – quite understandably – cannot accept that there can be a fundamental discontinuity in reality that irreconcilably divides it into a physical reality containing physical states and a mental reality containing mental states. In general then, materialists begin by claiming that all things must *somehow* be part of a single undivided physical reality. But they then face the severe dilemma of trying to explain how this can be so, given that people clearly do have direct subjective experiences of mental events, such as emotions and qualia, and genuinely cannot conceive of how ‘what it is like’ for them to have such direct experiences of these things could ever be described in terms of ‘physical states’ arising in the phenomenal space which, under the TWP, is presumed to hold all of the contents of physical reality.

The way the conceptual framework developed in this note untangles the materialist’s dilemma is to suspend the assumptions underlying the TWP and accept that W[r] and a person’s W[i] are *not* the same thing, and should *not* be held to be in identity. As a corrective, physical reality can then – as already proposed above – be held to consist *only* of the contents of W[r], the contents of a person’s W[i] can then be held to be generated by information processing systems in their brain[r], and the contents of their M can be held to be generated by separate, if connected, information processing systems that are also in their brain[r].

If a person’s phenomenal space and its contents – all thing[i]s such as rock[i]s, bullet[i]s and brain[i]s – are by this means released from having to be the same thing as physical reality, then the fact that the contents of their M do not exist in that phenomenal space, and cannot be reduced to any ‘physical state’ supposedly arising in that phenomenal space, creates *no* discontinuity in the physical reality of W[r].

Moreover, the relationship between ‘physical states’ subjectively experienced as phenomena in the phenomenal space of a W[i] and ‘mental states’ subjectively experienced as non-spatial ‘phenomena’ ‘outside’ phenomenal space can be seen simply as a relationship between two subjectively experienced sets of phenomena, where the contents of both those sets – as well as the subjective character of ‘what it is like’ to experience their contents – will *all* supervene on contents of the single and *only* physical reality, W[r]. Specifically, they will all supervene on contents of W[r] that include the evolving states in a person’s brain[r] that arise as parts of that brain[r] receive and process information from, and express information into, other parts of that brain[r], its immediate environment of B[r], and its wider environment of W[r].

While this formulation may at first seem unfamiliar, it answers question 3(a) above and with this sufficiently corrects the ill-posed questions 3(b) and 3(c) to allow both broadly to be answered with a ‘no’.

---

<sup>81</sup> See Section 1, and also consider the claims made by proponents of the predictive processing and phenomenal self-model paradigms about how a person’s phenomenal experience of the physical world is generated.

<sup>82</sup> Noting that with this, in implementing the TWP people will also tacitly assume that their phenomenal space – and all of its contents – is in identity with all of the respective phenomenal spaces and their contents that are being experienced by all other people, such that all TWP implementers – i.e. all normal people – will be in ongoing tacit agreement that the contents of this presumed ‘single’, shared phenomenal space, including the presumed single space itself, is what constitutes physical reality. (Precisely as per Section 3(α) and (β) above.)

<sup>83</sup> For perspectives on dualisms and materialisms as they relate to the Mind Body Problem see (39).

## 8. Remarks on ‘consciousness’

A central question for philosophy of mind and related disciplines is whether, and if so how, a person’s subjective experience – their consciousness – of being a self-in-a-world can be generated through the operation of information processing systems that exist and function solely in the physical world.

The conceptual framework developed in this note provides an understanding of what is meant when people refer to such a ‘physical world’. One of the framework’s key features is that it allows adoption of the position that what it defines as the noumenal world,  $W[r]$ ,<sup>84</sup> can be held to be the *only* physical reality, that all information processing will take place solely in this physical reality, and that everything a person experiences as phenomena, thinks and does, including all of their communications, will supervene on states,<sup>85</sup> and the evolution of such states, that exist solely within that physical reality.

Consistent with this, it has been proposed above that at the level of physical reality, defined as the noumenal world,  $W[r]$ , a person exists only as a noumenal body,  $B[r]$  – housing their noumenal brain[r] – in interaction with the rest of  $W[r]$ . One way to express this level of existence for a person is to call it their  $\{B[r] \text{ in } W[r]\}$ .<sup>86</sup>

At the same time, it has been proposed above that, through information processing in a person’s brain[r], a person’s subjective experience – i.e. their direct perceptual experience – of themselves as a physical being existing in, and interacting with, a physical world is an experience of being a phenomenal body,  $B[i]$ , immersed in a phenomenal world,  $W[i]$ . One way to express this level of existence for a person is to call it their  $\{B[i] \text{ in } W[i]\}$ .<sup>87</sup>

So with this, it can be proposed that we perceive ourselves, and we are *conscious* of ourselves as being a  $B[i]$  existing in, and acting on, a  $W[i]$ , which we call “the physical world” – i.e. we are conscious of ourselves as being a  $\{B[i] \text{ in } W[i]\}$  – whereas the physical reality is that we are a  $B[r]$  existing in and acting upon  $W[r]$  – i.e. we are a  $\{B[r] \text{ in } W[r]\}$  – and that all of the information processing underpinning the generation of the consciousness referred to is taking place in  $W[r]$ .

The question then arises as to what kind of information processing system resident and operating in  $W[r]$  – most likely within a person’s brain[r] housed within their  $B[r]$  – could generate consciousness?

One way to seek to answer this question is to begin by noting that, as described in Section 1, the information processing system  $D[r]$  does more than just take sensory input,  $s$ , caused by the evolving “hidden states” of a person’s  $\{B[r] \text{ in } W[r]\}$  and process this, by implementing an RBC, to generate that which the person experiences as an evolution of phenomenal states,  $\{B[i] \text{ in } W[i]\}$ , that maps those evolving “hidden states”. A person’s  $D[r]$  – as it sits within their brain[r] – *also* generates motor output that is expressed by a person’s  $B[r]$  as it acts both upon itself and on its wider environment in  $W[r]$ .

Say then, that by some means, information processing and storage in  $D[r]$  proceeds in such a way that, as a person grows from infancy, systems within their  $D[r]$  ‘learn’ by trial and error – the trial and error evident in an infant’s motor output as it progressively learns to coordinate movement of its  $B[i]$ <sup>88</sup> – a ‘vocabulary’ of motor output ‘phrases’ such that:

- when a fully physically coordinated adult phenomenally experiences some initial state,  $\{B[i] \text{ in } W[i]\}^a$ ,
- and envisages – and then seeks to move to – a phenomenal experience of a different state,  $\{B[i] \text{ in } W[i]\}^b$
- their  $D[r]$  will have learnt to – and will – spontaneously generate and *reliably* express the right motor output ‘phrases’ needed to move the state of the person’s  $\{B[r] \text{ in } W[r]\}$  from  $\{B[r] \text{ in } W[r]\}^a$  to  $\{B[r] \text{ in } W[r]\}^b$
- where, due to that expression, the person will phenomenally experience themselves as moving towards – and then as attaining – the envisaged phenomenal state  $\{B[i] \text{ in } W[i]\}^b$ .<sup>89</sup>

The last bullet point will be so, since:

- $\{B[i] \text{ in } W[i]\}^a$  will map  $\{B[r] \text{ in } W[r]\}^a$ ;
- $\{B[i] \text{ in } W[i]\}^b$  will map  $\{B[r] \text{ in } W[r]\}^b$ ; and so
- the transition from  $\{B[i] \text{ in } W[i]\}^a$  to  $\{B[i] \text{ in } W[i]\}^b$  will map the transition from  $\{B[r] \text{ in } W[r]\}^a$  to  $\{B[r] \text{ in } W[r]\}^b$ .

<sup>84</sup> See Section 1, *noting well* that the definition of the term ‘noumenal world’ as it is used throughout this note is subject to the caveat at footnote 4.

<sup>85</sup> Where these can be considered equivalent to the “hidden states” of the predictive processing paradigm.

<sup>86</sup> Where this is simply a shorthand expression for a person’s disposition as the evolving state of their noumenal body in the noumenal world.

<sup>87</sup> Where this is simply a shorthand expression for a person’s *perception of* their disposition as the evolving state of their phenomenal body in a phenomenal world, where their  $\{B[i] \text{ in } W[i]\}$  will *map* part of the evolving state of their  $\{B[r] \text{ in } W[r]\}$ .

<sup>88</sup> Noting that this ‘learning’ is not – at least to begin with – a conscious process, and that an infant’s degree of awareness/consciousness of itself as a physical body grows from near zero along side growth in its degree of competency in sensory-motor coordination.

<sup>89</sup> This four bullet formulation can be equated to ‘active inference’, as described in the predicative processing paradigm (29).



With this in mind, it can be proposed that a person's conscious experience of being a physical body immersed in and interacting with a physical world – i.e. their conscious phenomenal experience of being a  $\{B[i] \text{ in } W[i]\}$  – will arise as, and when, their  $D[r]$  processes information *with respect to itself* as if it actually *is*, or is 'somewhere within', the person's  $B[i]$ .

Put slightly differently, the proposal here is that if a person's  $D[r]$  comes to utilize some *mechanism* in its information processing operations that places it into *identity* with a *representation of itself* – i.e. places it into identity with, or 'within', the  $B[i]$  that it is itself generating – then for periods over which that mechanism is engaged the person's  $D[r]$  will have a conscious phenomenal experience of itself as being a  $\{B[i] \text{ in } W[i]\}$ .

This proposal aligns with how we consciously experience ourselves as physical beings in a physical world. We experience ourselves as either being – or as being somehow 'within' – our  $B[i]$ , and as moving our  $B[i]$  such that we are able to move from being in some given state,  $\{B[i] \text{ in } W[i]\}^a$ , to some firstly envisaged and subsequently realized state,  $\{B[i] \text{ in } W[i]\}^b$ .

If so, the question then becomes how to describe a mechanism by which a processor  $D[r]$ , within a person's  $B[r]$ , can come to sustain – and thereby operate as if – it is in identity with its own representation of itself,  $B[i]$ .

One candidate mechanism can broadly be described as follows.

When a person's  $D[r]$  – which is an *entity[r]* – is moving one given "hidden state"  $\{B[r] \text{ in } W[r]\}^a$  to any other "hidden state"  $\{B[r] \text{ in } W[r]\}^b$  it will – *iff* it has 'learned' the right motor output 'phrases' to \*high reliability\* – be able to operate on a presumption that it *is* in fact an *entity[i]* which it will be able to *infer* with commensurately \*high reliability\* through 'observing' the outcomes of its processing of sensory input, is precisely *that* *entity[i]* which is moving  $\{B[i] \text{ in } W[i]\}^a$  to  $\{B[i] \text{ in } W[i]\}^b$  where, *to best fit*, that inferred *entity[i]* must be either  $B[i]$  itself, or somehow be 'within' and driving the motions of  $B[i]$ .

That said, it is not within the purview of this note to further detail this kind of approach to solving how a person's 'consciousness' of being a physical being in a physical world might be generated.<sup>90</sup>

Rather, the key point to be made here is that if the working assumptions of the TWP are held to be objective truths rather than assumptions – meaning that what have been described above as  $W[r]$ ,  $W[i]$  (and  $W[z]$ ) must for all purposes be held to be the same thing – i.e. be held to be in identity with each other – then this would preclude development of approaches of the kind just outlined to solving how 'consciousness' and its accompanying subjective states might be generated.

This is because if we each consider it an objective truth – rather than a *tacit assumption* under the TWP – that our  $W[i]$  is in identity with  $W[r]$  – and with this, that our  $\{B[i] \text{ in } W[i]\}$  is in identity with our  $\{B[r] \text{ in } W[r]\}$  – then this removes all room for any proposal that some form of mechanism – such as the candidate mechanism for generating consciousness described above – can exist to enable that tacit assumption to be sustained.

More importantly, if it is considered an objective truth that a person's  $W[i]$  is in identity with  $W[r]$ , this would preclude all predictive processing and phenomenal self-model accounts of how phenomenal experience may be generated since – as shown in Section 7 – both of those accounts rest on a distinction between  $W[r]$  and  $W[i]$ . In turn, this would also preclude any approach to solving the problem of 'consciousness' based on either predictive processing or phenomenal self-model accounts of phenomenal experience.

For these reasons, as well as the reasons provided in Section 7, it is the summary contention of this note that the objective reality is that:

- the  $W[i]$  that any person experiences is not, never has been, and never will be, in identity with  $W[r]$ ;
- the phenomenal reality of a person's perception of themselves as a  $\{B[i] \text{ in } W[i]\}$  is not, never has been, and never will be, in identity with the physical reality of their existence as a  $\{B[r] \text{ in } W[r]\}$ ;

yet that

- under the TWP, it is a biologically natural, highly successful, and evolutionarily driven adaptation for human beings *always* to operate both individually and collectively on a tacit assumption – i.e. *always* to operate *as if* – the contents of our respective  $W[i]$ s are in identity with contents of  $W[r]$  and, for virtually all purposes, are also in identity with contents of the  $W[z]$  of our language using group.<sup>91</sup>

<sup>90</sup> Noting that the author has extensively explored this type of approach elsewhere (40).

<sup>91</sup> Where this last point can be considered an elucidation of the naïve realism that we *do* each universally apply as we go about our everyday lives, including as, amongst ourselves, we communicate, cooperate and coordinate our physical actions to improve our physical circumstances.

## Bibliography

1. *Immanuel Kant's Critique of Pure Reason* trans. Kemp Smith, N. Macmillan & Co., London, UK (1929) – in particular the *Transcendental Aesthetic*.
2. Koch, C. (2018) What is Consciousness? *Nature* 557 S9-S12.
3. Metzinger, T. *Being No One: The Self-Model Theory of Subjectivity* MIT Press, Cambridge MA, USA (2003).
4. Metzinger, T. *The Ego Tunnel: The Science of the Mind and the Myth of the Self* Basic Books, New York (2009).
5. *Ibid.* pp 107.
6. Ismael, J.T. *The Situated Self* Oxford University Press, New York, USA (2007).
7. *Ibid.* pp 229-230.
8. Hohwy, J. *The Predictive Mind* Oxford University Press, New York, USA (2013).
9. *Ibid.* pp 227-230.
10. Clark, A. *Surfing Uncertainty: Prediction, Action and the Embodied Mind* Oxford University Press, New York, USA (2016).
11. *Ibid.* pp 195-196.
12. *Ibid.* pp 170-171, 194-196.
13. Friston, K.J. (2003) Learning and Inference in the Brain. *Neural Networks* 16 1325-1352.
14. Friston, K.J. (2010) The Free-Energy Principle: A Unified Brain Theory? *Nature Reviews Neuroscience* 11 127-138.
15. Wallace, D. and Carroll, S.M. in conversation at 11.15 to 13 minutes in *Mindscape 158* at <https://www.youtube.com/watch?v=ZusCQdM6ovw>.
16. See esp. Section 2 in Faye, J. (2019) Copenhagen Interpretation of Quantum Mechanics *The Stanford Encyclopedia of Philosophy* (Winter 2019 Edition) Zalta, E.N. (ed.) <https://plato.stanford.edu/archives/win2019/entries/qm-copenhagen>.
17. Wallace, D. (2012) Decoherence and its Role in the Modern Measurement Problem *Phil. Trans. R. Soc. A* 370 4576–4593.
18. Popper, K.R. (1978) *Three Worlds* The Tanner Lecture on Human Values, University of Michigan. [https://tannerlectures.utah.edu/\\_resources/documents/a-to-z/p/popper80.pdf](https://tannerlectures.utah.edu/_resources/documents/a-to-z/p/popper80.pdf).
19. Popper, K.R. On the Theory of Objective Mind, Ch.4 in *Objective Knowledge* Rev Ed. Oxford University Press, New York, USA (1979).
20. (a) 'Mary's Room' as originally formulated, is provided in Jackson, F. (1982) Epiphenomenal Qualia *Philosophical Quarterly* 32 127–136.  
(b) For wider context see Tye, M. Qualia *The Stanford Encyclopedia of Philosophy* (Fall 2021 Edition), Zalta, E.N. (ed.) <https://plato.stanford.edu/archives/fall2021/entries/qualia> and references therein.
21. Champagne, M. (2019) Consciousness and the Philosophy of Signs *The American Journal of Semiotics* 35 443-462.
22. Block, N. (1990) Inverted Earth *Philosophical Perspectives* 4 53–79.
23. Armstrong, K. *A History of God: The 4,000-Year Quest of Judaism, Christianity and Islam* Ballantine Books, New York, USA (1993).
24. Papineau, D. (2000) The Rise of Physicalism, Ch.10 in *Proper Ambition of Science* Eds Stone, M.W.F. & Wolff, J., Routledge <https://doi.org/10.4324/9780203446263>.
25. Definitions in relation to physicalism are drawn from Stoljar, D. (2022) Physicalism *The Stanford Encyclopedia of Philosophy* (Summer 2022 Edition) Zalta, E.N. (ed.) <https://plato.stanford.edu/entries/physicalism/> and references therein.
26. Ramstead, M.J.D. et al. (2020) A Tale of Two Densities: Active Inference is Enactive Inference *Adaptive Behaviour* 28 225–239.
27. Friston, K.J. et al. (2012) Perceptions as Hypotheses: Saccades as Experiments *Frontiers in Psychology* May 2012/Volume 3/Article 151; doi: 10.3389/fpsyg.2012.00151
28. Parr, T & Friston, K.J. (2018) The Anatomy of Inference: Generative Models and Brain Structure *Frontiers in Computational Neuroscience* 12:90; doi: 10.3389/fncom.2018.00090
29. Hohwy, J. (2020) New Directions in Predictive Processing *Mind & Language* 35 209-223.
30. Smith, R. et al. (2020) An Active Inference Approach to Modeling Structure Learning: Concept Learning as an Example Case *Frontiers in Computational Neuroscience* 14:41; doi: 10.3389/fncom.2020.00041
31. Friston, K.J. et al. (2020) Sentience and the Origins of Consciousness *Entropy* 22 516; doi: 10.3390/e22050516
32. Revonsuo, A. *Inner Presence: Consciousness as a Biological Phenomenon* MIT Press, Cambridge MA, USA (2006).
33. *Ibid.* pp 92-93.
34. *Ibid.* pp 181-195.
35. Seth, A. *Being You* Faber & Faber, London, UK (2021) pp 75-76.
36. Ismael, J. (2020) What Entanglement Might Be Telling Us: Space, Quantum Mechanics, and Bohm's Fish Tank, Chapter 7 in *The Foundation of Reality: Fundamentality, Space, and Time*. Eds Glick, D., Darby, G. & Marmodoro, A. Oxford University Press. <https://doi.org/10.1093/oso/9780198831501.003.0008> Or see: [https://www.researchgate.net/publication/341546733\\_What\\_Entanglement\\_Might\\_Be\\_Telling\\_Us\\_Space\\_Quantum\\_Mechanics\\_and\\_Bohm's\\_Fish\\_Tank](https://www.researchgate.net/publication/341546733_What_Entanglement_Might_Be_Telling_Us_Space_Quantum_Mechanics_and_Bohm's_Fish_Tank)
37. Ismael, J. (2021) Do You See Space? How to Recover the Visible and Tangible Reality of Space (Without Space), Chapter 9 in *Philosophy Beyond Spacetime: Implications from Quantum Gravity*. Eds Wüthrich, G., Le Bihan, B. & Huggett, N. Oxford University Press. <https://doi.org/10.1093/oso/9780198844143.003.0009> Or see: <https://www.jenanni.com/papers/DoyouSeeSpace.pdf>
38. Mermin, D.N. (2018) Making Better Sense of Quantum Mechanics *Reports on Progress in Physics* 82 012002.
39. See Robinson, H. (2020) Dualism *The Stanford Encyclopedia of Philosophy* (Fall 2020 Edition) Zalta, E.N. (ed) <https://plato.stanford.edu/archives/fall2020/entries/dualism/>.
40. Hammer, B.C. <https://www.researchgate.net/profile/Brendon-Hammer> see featured research and, under research, Working Note A.