Running into Consciousness

John Barnden
School of Computer Science
University of Birmingham
Birmingham B15 2TT, U.K.
J.A.Barnden@cs.bham.ac.uk

Abstract

It is proposed that conscious qualia arise when and only when the 'running' of physical processes takes a special, complex form. Running in general is the unified unfolding of processes through time, and is claimed to be an additional quality of physical processes beyond their state trajectories. The type of running needed for conscious qualia is reflexive in physically affecting and responding to itself. Intuitively, running is essentially the flow of causation, and the self-affecting/responding is a matter of causation bearing a causal relationship to itself: that is, causation can itself be reflexive. The proposal potentially makes temporal qualia the most fundamental qualia, with others derivatively arising. The proposal is neutral about whether a conscious process occurs in a natural organism, in a physically implemented computational system, or in some other physical substrate, as long as the substrate involves the special reflexive runningness.

1 Introduction

This article makes a proposal about the "hard" problem of consciousness: the nature of phenomenal consciousness. I will just say "consciousness" to mean phenomenal consciousness: consciousness that involves qualia (conscious feelings, sensations, emotions, imaginings, etc.). The problem is the fact that it feels like anything at all to be awake as opposed to sleeping dreamlessly. Why and how can that sort of consciousness exist? How can it arise out of physical matter?

I will use the verb "feel" to cover the having of any sort of quale. So the quale of seeing blue is a feeling, as is the quale of hearing a particular sound, or having a pain, experiencing an emotion, engaging in imagery (visual, auditory, kinaesthetic, ...), or feeling that time is passing. I will use the nouns "feeling" and "quale" interchangeably, so that feelings in this paper will always be conscious ones. I also include in feeling the qualia, if any, involved in thinking consciously about something, or having an occurrent belief, intention, etc. Thus, feelings and qualia in this paper are very broad categories, not confined to emotion, sensation and sensation-like imagery.

The proposal will centre on three bold suggestions, which in intuitive causal terms can be expressed as follows:

- That physical causation can itself take part in causation, and in particular an instance of causation can be in a causation relationship to itself (reflexive causation). That is, there are cases of A being in a causal relationship with B where A and/or B are themselves instances of causation. And, more specially, there are cases of this where B is A, so A is in a causal relationship with itself.
- Some instances of reflexive causation are also instances of occurrence of a very basic quale (conscious feeling). This quale is referred to below as the *fundamental quale*. Notice that the suggestion is a metaphysical one about what that quale *is*: the quale doesn't simply accompany suitable types of reflexive causation—it *is* an aspect of such causation.
- All qualia are somehow founded on the fundamental quale, via (in the human case) multiple layers of complex physical interactions in the brain, some of which can be described as cognitive operations such as thinking.

I will soon dispense with the notion of causation except for heuristic purposes of exposition. But first: the overall structure of the argument in this paper will be as follows. I will present evidence that reflexive causation is a *necessary* aspect of consciousness. I will then (merely) suggest that it is *sufficient* for consciousness, via the above suggestion about the fundamental quale. I present no direct argument that it is sufficient, but will argue that the suggestion has certain beneficial consequences, aiding the analysis of some long-standing issues. Hence, it should be considered as a serious competitor to other proposals about consciousness.

It seems that the paper's proposal has not been made before, at least in quite such an explicit form, although there are some strongly related proposals. The paper is therefore intended as a preliminary placing of an idea on the stage of debate. For this reason it is useful briefly to analyse a variety of ramifications of the idea, rather than developing any one ramification in great depth.

This paper makes the working assumption that consciousness is a fact about the physical universe. Qualia are real aspects of physical reality and not epiphenomenal: they engage with other aspects of reality via (ultimately) physical laws—consciousness can have physical effects just as any other physical quality can. More precisely, not to prejudge too closely what the "physical" is, consciousness is neither a purely abstract entity akin to a mathematical structure nor an aspect of some separate, concrete realm of being, such as a spiritual realm. It is difficult to put this point more clearly, as there is much unclarity and contention about what the "physical" is (cf. Davies, this issue). I adopt a physicalist position partly to avoid multiplying realms of being unnecessarily. It's important to see whether we can explain consciousness without supposing a different order of being such as a spiritual realm.

The notion of causation was used above for expository convenience. However, it is an insecure basis for a discussion of consciousness as emerging from the physical world (cf. Leech Anderson 2012). Causation is not a technical part of physics, and is beset by many philosophical problems. Therefore, the paper assumes merely that nature (the universe) operates by physical laws. The notion that X causes Y is arguably just a heuristic abstraction from X and Y being somehow related by the action of physical laws.

I will say that a physical process—the temporal evolution of a part of the universe over some short or long period of time as governed by physical laws—involves *running*, where running is the operation of physical laws over a duration in some portion of the universe. It is an essential aspect of physical processes. Running here is conceived of as something that, metaphorically speaking, connects the physical states or subprocesses of the process together. Crucially, *a physical process is not just its state trajectory—the amalgamation of its instantaneous physical states:* rather, it also *runs*. Thus, running is itself a real feature of the physical universe, and perhaps *the* real thing.

Given that running is itself physically real, and goes beyond the process's state trajectory, it is not a large step to conjecture that the running inherent in at least some physical processes can have its own physical effects, beyond the overall effect of the state trajectory alone. More precisely, runningness itself can take part in physical laws, both as a determining variable and as a determined one—albeit laws of a type different from those already existing in physics. That is, if you somehow achieved the impossible feat of artificially arranging a continuous sequence of instantaneous states (each being a particular arrangement of mass, energy, momentum, velocity, acceleration, electrical and magnetic fields, etc. in particular spatial configurations) that copied the states of the physical process, but that did not naturally unfold as that process does, there could be different physical effects from those of the original process.

The operation of physical laws is, intuitively, causation, so an instance R of running is also, intuitively, causation. So we are intuitively just saying that *causation itself* can take part in causal relationships, which could be called meta-causation relationships. Alternatively, in the non-causal terms of runningness, we have it that instances of running, not just physical states connected by running, can themselves be connected by running, and the latter running could be called meta-running.

The paper will also claim, as already presaged, that there is *reflexively-acting* running, which is a matter of an instance R of running responding to and affecting itself. It will be assumed that this reflexively-acting running, although a special type of meta-running, is part of the running instance R itself (not something outside or on top of it). Thus the reflexive action of R is an intrinsic constituent of R. So, intuitively again, a particular case of causation can be in part *internally self-causing,* not just entering into causal relationships with other things outside itself as in generic meta-causation.

Given the above, I then just postulate that at least some reflexively-acting running just is (or is partly constituted by) a fundamental quale, or perhaps a set of qualia. Thus, this quale is a part of the very tissue of portions of the physical universe. This fundamental quale is of a very basic, undifferentiated, non-cognitive kind, and may amount just to a *feeling of existing,* with an inherent reflexive quality of feeling-of-being-a-feeling.

The paper is not proposing here a separate subject that feels the quale, although it is probably appropriate to say the quale reflexively feels itself (in some very basic way that does not involve construing a "self"). The notion of a feeling subject is, in this paper's view, derivative from qualia, when of a suitable sort and suitably arranged, with cognitive elements such as reasoning thrown in. A notion of subject is not needed as part of the concept of qualia. Also, notice that the fundamental quale is not something separate from the running. It *is* the special sort of reflexively-acting running, or at least an aspect of it.

As a sub-proposal I will more tentatively conjecture that the quale is a feeling of continuation (time passing). This then connects in an important way with ongoing controversies about the nature of temporal consciousness.

In its building in of a quale or some set of qualia into the fundamental fabric of the universe, the proposal is reminiscent of panpsychic approaches to consciousness (Skrbina 2009, Weekes 2012), and is also a form of monism. However, it is not proposed that all running is reflexively-acting, but just that there is some form of it, in some portions of the universe (e.g., somewhere sometimes in some brains), that is reflexively-acting. So there is no necessary step into panpsychism, which is fortunate if panpsychism is generally viewed as implausible (see, e.g., Bishop 2009). Most running may be perfectly ordinary and non-experiential/qualiaful. What I'm saying is that conscious entities are based, somehow and somewhere in their constitution, on a special case of running: some suitable form of reflexively-acting running. I would therefore say that the

proposal is a *bathypsychic* one (with "bathy" conveying "deep" as in "bathysphere", a vessel that goes down into the deep), rather than a panpsychic one.

The proposal in this paper points, I will argue, towards fresh approaches to some long-standing issues about consciousness. First, it helps with the question of what sort of physical realization, if any, is needed for consciousness. The proposal subscribes to a wild multiple realizability: there is (presently) no reason to think there is a limit to the specific physical substrates on which conscious computation can in principle be realized (human brains, computers, fluidic circuits, alien non-carbon-based brains, ...). However, *some* physical realization is needed. and must rest on the above special form of runningness. This does not preclude certain substrates being more suitable in practice. For instance, the special runningness may, conceivably, benefit from a certain type of cognitive organization. Perhaps, say, a global workspace structure (Baars 1988; Dehaene & Naccache 2000) in practice facilitates it. Secondly, the proposal illuminates what it would take for an artefactual computational system to be conscious. Thirdly, the proposal points towards an alternative, advantageous way of bringing together two seemingly unrelated aspects of human consciousness: basic qualia such as pain, and higher-order thinking. It also brings to the fore and provides improved grounding for a common phenomenal aspect of different qualia (of vision, hearing, imagining, emoting, thinking, or whatever), despite their felt difference. Finally, it potentially helps with problems about temporal consciousness.

The proposal relates strongly to certain important movements in philosophy and elsewhere. First, it is an instance of "process philosophy" (Seibt, 2013; Weber & Weekes, 2010; Weekes 2012), and in particular takes on board Whitehead's insistence on events being the foundation of the universe, with persistent objects needing to be constructed from events (Whitehead 1929/1978, 1984; Weekes 2012). Also, via runningness's intimate connection to time, it resonates with widespread claims that temporality is central to consciousness (e.g., Husserl 1991, Lloyd 2012, Nunn 2010, 2013; and see Dainton's 2014 survey).

The structure of the paper is as follows. Section 2 describes the proposal and the progression of ideas involved in it. Section 3 adds further discussion, concentrating on ways in which the proposal may help us. Section 4 introduces a sub-proposal, that the fundamental quale is temporal. Section 5 makes some final remarks.

2 The Proposal

The Introduction laid out some of the ideas and the shape of the connecting argument. This section lays them out more carefully, starting from scratch (i.e. not simply assuming the truth of the claims made in the Introduction).

2.1 Some Initial Assumptions

The proposal starts from two initial assumptions, both of which are intended to be uncontroversial and reflective of general views in the field. First, consciousness is a (physical) process of a particular sort, or at least an aspect of processes of a suitable sort. To put it another way, consciousness is a matter of *activity over time* in the world. I do not assume that the process has to be a continuous one. It's possible that my proposal leads to a conjecture that it must be continuous, but this is not crucial to the argument in this paper.

The second initial assumption is that consciousness is inherently *reflexively-acting*–consciousness essentially involves self-responding and self-affecting. To assume this does *not* just build in the core suggestion of this paper, namely that there is a quale that is a type of reflexively-acting causation (running). Rather, the assumption is just an observation of a key feature of phenomenal consciousness, *whatever* its nature is. We have not yet, in this section, arrived at the question of running.

Also, the reflexive action does not entail that a conscious process has complex cognition, based for instance on conceptualizations about itself, or that there is a sense of a "self" in the consciousness. The reflexively-acting is intended to include a primitive sense in which a conscious process *feels itself to exist*, rather than necessarily *thinks that* it exists. Furthermore, the assumption is to be understood as requiring that consciousness responds to and has effects on *itself*, not just to/on some *representation of* itself (although it may do this too).

2.2 The Main Progression of Ideas

The progression of ideas is organized into the Steps below. In outline, these steps are as follows. Step 1 will develop the idea of reflexivity, and uses it to suggest that a conscious process must respond to its own running. Steps 2 and 3 will infer from this that meta-causation and indeed reflexive causation (meta-running and reflexively-acting running) must exist. So far there is no postulate about what qualia are: the argument has merely provided evidence that reflexively-acting running is *necessary* for consciousness, and not that it is *sufficient* for providing qualia. But Step 4 adds the suggestion that there is a fundamental quale that is a type of such running. Step 5 raises the so-called "combination" problem of how complex, high-level qualia might be based on a fundamental quale.

2.2.1 Step 1:

According to the reflexivity assumption above, the conscious process responds to *itself*. I claim here that it is responding to itself *as a process*, otherwise it is not truly responding to its own reality. But this has a fundamental consequence, given that, as noted above, no process is just its state trajectory: it also crucially involves connectivity between the trajectory's states. This connectivity is runningness, or, intuitively, causality.

The fundamental consequence is that *the reflexively-acting is not fully characterizable in terms of the trajectory as such:* a full characterization must bring in the runningness. The way the process affects itself is not reducible to the way that the states in the state trajectory depend on states in the state trajectory.

In brief, it matters to any conscious process that it is indeed a *process*. I don't mean by this that the consciousness necessarily has a *concept* of process (although human level consciousness may do) but merely that something about the reflexively-acting rests crucially on the fact that what is being responded to and is being affected is a process. This point about mattering-to-itself-as-process can be seen in work on mind over a long period, for instance discernible in von Foerster (1976) and explicit in Fekete & Edelman (2011). However, such works do not draw out the consequences discussed in the present paper.

To some extent it is just a postulate that consciousness responds to and affects itself *as a process*, rather than consciousness just (a) being a process that (b) matters to itself in some other way. However, there

is an argument that one can bring forward in support of the idea that the processuality is essential to the self-mattering (i.e., to the reflexively-acting).

Suppose that the reflexively-acting needed for consciousness could be characterized entirely in terms of state trajectories. Imagine a conscious process C, where we restrict attention to the portion of the process up to some time *t*. By the supposition, all that matters in this portion is the state trajectory, and the way states in this trajectory have somehow affected other states, in a way describable overall as reflexively-acting. The causality/runningness across the states is irrelevant. The argument then is that we lose nothing by "staticizing" the process portion up to *t* into a timeless, purely spatial version in the following way (cf. related observations in Schweizer 2014; also the argument is similar in spirit to many thought experiments in the literature—see for instance Bishop 2009). To whatever level of accuracy one wished for, one could chop up the state trajectory of the whole process up to time *t* into a finite succession of time instants (or perhaps the original process was already defined over a finite, discrete sequence of instants). We can then imagine the states of the system at these instants being realized in some way, simultaneously, in separate pieces of physical stuff. One would have a static arrangement of chunks of stuff, each of which copies an instantaneous state of the original system, but existing simultaneously. (An example of this would be if the physical system in question was a running computer, and each simultaneous chunk was an exactly similar computer, each statically holding one state of the original computer.) The chunks need have no physical connection to each other. Nevertheless, the structure of all the interactions in the original system, including the reflexively-acting, would be preserved, to whatever level of accuracy required, by making the chopping up suitably fine. But it would seem remarkable to maintain that the resulting set of chunks would still be conscious, or more precisely, would be serving as a substrate of an overall physical state that is conscious.

Now, one might complain that we have lost any progression through time: the thought experiment isn't fair, because the states are no longer in temporal sequence. One no longer has any reflexively-act*ing*, but just a copy of the *structure* of the reflexively-acting in the original process. But one could do something like successively marking each of the chunks by some marker (e.g. shining a light on chunks successively). We now have a system that can be construed as running through the allegedly appropriate sequence of states (namely, the highlighted portions of the state of the whole set of chunks). This sequence includes all the reflexively-act*ing*, unfolding through time just as it did in the original system. Or, if one is uneasy that the question should rest so much on an act of construal, one could imagine introducing the right state into each chunk successively, rather than having them present simultaneously. But why should it matter, to the question of the chunk-set possessing consciousness, whether the individual chunks get their states in sequence or hold them simultaneously, especially if there is no physical relationship between the chunks at all?

The central point is that if one claims that the runningness (causality) between the states is irrelevant, it is difficult to see why *time* itself should matter. What significance does it have to the individual states what time sequence they occur in, or that they occur in *any* sequence? Note here that in the staticization exercise one could add to the states all needed information about what (relative) times they occurred at in the original state trajectory, so no relative-time information is lost.

This argument would be negated if it were justifiably maintained that the structure of the necessary reflexively-acting would not be preserved by such a staticization, no matter how finely one chopped, because e.g. of chaotic effects that are essential to the original process. So I put the argument forward as merely suggestive of the point that it is possible that we are not confined to postulating baldly that a conscious process needs to reflexively act upon itself as a process. Rather, that it needs to do this may follow from principled considerations.

This Step allows that an episode of consciousness might *also* be accompanied by processing that does just react to its own state trajectories as such, ignoring runningness. But that reflexivity would in itself be of an unconscious sort, because it cannot respond to itself as a *process* complete with its runningness (its causal linkage). Not all reflexivity amounts to consciousness.

2.2.2 Step 2:

But now note an important implication of Step 1. I'll first state it in the intuitive language of causality. The implication is that *causation itself can cause and be caused,* or more precisely there can be a causal relationship between something A and something B where an aspect of A and/or an aspect of B is itself an instance of causation. For, surely, the reflexively-acting is itself a species of causation. In order for the causal flow in the conscious process to matter to that conscious process, the causal flow must, as part of the reflexively-acting, have causal effects on itself.

This contrasts with an ordinary causal flow, as for instance when the throwing of a ball causes a window to break. This is not a matter of causation causing anything, let alone aspects of itself. It is events or states that are joined in the flow that cause each other—it is not that some part of the *causal flow* causes later parts. The point of my argument is that the flow that ties together the constituent events and states of a conscious process, and is separate from those events and states, itself has causal power and susceptibility.

I have used causation as intuitive clothing for the notion of nature unfolding via physical laws. In non-causal terms, what the above means is that, within a process that is a consciousness, the running (i.e., the unfolding of nature through time, via physical laws) is itself something that takes part in physical laws. Further, this taking part is its reflexive responding to and acting upon itself. Thus, the laws are of a new type that relate at least in part to running as such, not just to the normal properties appearing in laws. Note that *laws* as such neither run nor take part in laws. Nature runs, via laws (or perhaps even: nature is overall a running, via laws); and it is the running of nature via laws that is itself being said to be something that can take part in laws. So, for example, the principle of energy conservation does not run; rather, nature runs in a way governed in part by the law (or perhaps better: there exists running that is governed in part by the law). Also, the principle of energy conservation does not take part in laws; rather, it is running, partly governed by the principle, that takes part in laws.

The claim of a new level of law that governs running as such is of course a large one.

2.2.3 Step 3:

Furthermore, the running of the laws referring to the runningness of a conscious process must be an aspect of the original process itself, simply because the reflexively-acting is postulated to be such an aspect. So the process is inherently one where some running of physical laws connects aspects of that same running. The running is in part the reflexively-acting of that same running. The process of reflexively-acting is not an extra process that is, so to speak, "on top of" the running that is engaging in the reflexively-acting.

Although the notion of reflexively-acting running is, of course, somewhat obscure and mysterious, I believe it is more acceptable than thinking that causation is real and that there is self-causing causation. Quite apart from questions about the physical reality of causation, or the philosophical coherence of the notion, we

have entrenched intuitive views about what causation is that may get in the way of even conceiving of the reflexive variety. For instance, we often appear to metaphorically conceptualize an instance of causation as an ordinary commonsense-world physical force exerted by the causing entity on the affected entity (cf. the force-dynamics approach of Talmy 1988). It is difficult to conceive of such a force-relationship between two entities as *itself* being a cause or something caused, that is, being an entity that is in a force relationship with something else (let alone itself).

2.2.4 Step 4:

Here is the most postulatory part of the proposal. It is the metaphysical claim that (some suitable form of) reflexively-acting running *is* just feeling in some fundamental form (or has feeling as a constituent aspect). All feeling in consciousness, at whatever level of complexity or sophistication, rests somehow on this fundamental feeling. I am not here proposing a subject that engages in the feeling, except in that the quale itself reflexively feels itself.

The fundamental quale envisaged could be described as something like "feeling something or other" where there is no sense of what sort of feeling it is or of what is having the feeling. Indeed, there is no *conceptualization* here of feeling or of anything else. I'm not proposing that the special, reflexive flow is itself equipped with cognitive powers.

A more elaborate postulate would be that there are several distinct varieties of fundamental feeling, each being a different variety of reflexively-acting running, but for simplicity the following will just just stick to one.

2.2.5 Step 5:

The proposal raises the same "combination problem" much discussed in relation to panpsychism (see, e.g., Weekes 2012). The problem is that of explaining how, say, human-level consciousness, with all its richness, arises out of the very deep, fundamental consciousness forming the tissue of (some of) the universe, without just building some such richness into that deep tissue. Weekes (2012) explains how a Whiteheadian account can solve this by appealing to (what I here roughly express as) complex information processing over time involving various forms of high-level conceptualization that is going on within the mental process. I propose that something like this must hold, although I don't yet have a detailed account. Perhaps something on the lines of the detailed mathematical account of of Fekete & Edelman (2011), where in effect state trajectories respond to themselves, would help. I assume that it is possible for the richness and intensity of feeling, and its variety, to be the greater the more that the basic reflexively-acting running gets bound up in complex structures of information and processing of information.

When the fundamental quale was said in Step 4 to be part of law-governed reflexively-acting running, the door was left open for those laws to engage ordinary aspects of physical reality as well. Suppose that those ordinary aspects can support, in some ordinary implementational way, information structures and processing of information structures. Then the quale—or rather, instances of occurrence of the quale in many parts of the overall system—could become attached to such structures and processing. For instance, if that processing amounts to processing of visual input, and to the maintenance of a representation of the person's own body, then the overall array of instances of the quale could become mutually structured by

virtue of the structure of the information processing. Quale instances could thereby indirectly interact in complex ways, forming a dynamic, complex whole. The quale-complex is now reacting to itself as a quale-complex maintained by a particular body that is processing a particular type of information. This is far from a solution of the combination problem, and may just be a more detailed specification of what the combination is in the case of the proposal in this paper, but may be a useful start.

The overall system's information and processing of it can be a matter of beliefs, intentions, and so on. There is nothing about processing of such propositional attitudes that *per se* that brings in consciousness. One can have entirely unconscious propositional attitudes and mental processing that connects and creates propositional attitudes. However, when that processing is done through physical processing that rests on the special, reflexively-acting running as above, then aspects of it can be conscious.

2.3 Type of Physical Matter

The only variety of consciousness we know of securely is human consciousness, although many of us may strongly suspect that some animals are (phenomenally) conscious. Some people may also postulate conscious deities, or may feel that they are acquainted with a transcendent consciousness. But if we ignore these additional possibilities, it is in principle possible that a conscious process can only reside in biological matter, for some reason. If so, there is something special about biological matter that allows the reflexively-acting runningness discussed above, and that cannot be achieved—at all or so easily—in matter organized in some different way.

However, I see no reason to make such a supposition. The fact that the only consciousness we securely know of is human consciousness may in part be because we're human and in part because so far only humans have the requisite cognitive apparatus to think creatively about things like consciousness. It would beg the question against the possibility of conscious non-biological artefacts to assume, at present, that consciousness has to be a process in biological matter. Therefore, I provisionally propose that it *is a fundamental feature of the universe that running can take a reflexively-acting form and that this potential is not confined to cases of matter being organized in any particular higher-level way such as the biological.*

2.4 Some Relationships to Other Work

This paper's proposal is in a similar vein to several other proposals. First, Dainton (2014) quotes Husserl (1991: 84, 88) as saying

> There is one, unique flow in consciousness in which both the unity of the tone in immanent time and the unity of the flow of consciousness itself become constituted at once. ... The flow of the consciousness that constitutes immanent time not only exists but is so remarkably and yet intelligibly fashioned that a self-appearance of the flow necessarily appears in it, and therefore the flow itself must necessarily be apprehensible in the flowing.

This seems to be getting at something like the reflexive runningness essential the the present paper. The proposal in this paper is also akin to the idea that fundamental events (actual occasions) in Whitehead's philosophy (Whitehead 1929/1978) are subjective experiencers in some primitive sense.

Primas (2007) insightfully discusses mind as being a special case of one style of description of the universe that uses a tensed form of time, involving a Now etc., with the other style using the untensed time used in contemporary physics (although, importantly, tensed time also appears in physics in the form of the initial conditions set up in experiments). This work gives a mathematically precise and sophisticated account of how a tensed description and an untensed description can be complementary and incompatible, while being able to be rigorously stitched together in an overall scientific account. However, the runningness involved in tensed time is not explicitly given a reflexive quality. Of course, time-flow in general is a huge issue in consciousness research, as borne out by Dainton's (2014) survey.

The proposal by Fekete & Edelman (2011) is in a somewhat similar vein to the present paper's in being thoroughly process-based, and indeed it rests on a point very much like Step 1 in section 2.2. However, it approaches qualia in a way couched entirely in terms of the trajectories of states in processes, something the Step 1 discussion rejects as inadequate.

In the proposal of Nunn (2010, 2013), basic qualia arise out a breaking of a time-related symmetry at a fundamental physical level. I understand him to say that there is a quale that consists of what-it-is-like to be, or to be involved in, this breaking. It is possible that Nunn's proposal is a specialized form of the core of my proposal. Nunn (2010) gives a highly specific, biochemical clothing to his proposal, whereas the ideas in the present paper's proposal are independent of any particular biochemical claim and rest on more general considerations.

The most closely related proposal appears to be that of Baer (2010). In this, consciousness involves a type of self-"explanation" that must refer to processes (cf. Step 1 again). Furthermore, the relevant processes are Whiteheadian fundamental events. Consciousness is what it's like to be a cycle of mental activity of the sort described in Baer's paper, and involves getting into the cycle and feeling time flow through oneself. Baer also states that (roughly speaking) awareness of self-existence in empty space is a basic type of consciousness. This may be similar to the fundamental quale/qualia that the present paper's proposal says come out of suitable reflexively-acting running.

2.5 The Computational Case

The considerations so far, when specialized to the case of conscious AI systems, lead to the conclusion that such a system would need to be suitably implemented in a physical substrate that involves reflexively-acting physical running. Note that this observation rests only on steps 1–3 in section 2.2, arguing for the *necessity* of such running. It does not rely on the postulate in Step 4 about such running being *sufficient* for providing a quale.

Some further remarks are useful here. Suppose there is a conscious physical process that consists of a physically realized running of a computer program. For simplicity, I will consider only a sequential program running on a single CPU, not some arrangement where there is genuine parallelism between computation streams. Notice first the simple but crucial point that we do require the program to be *running*, and that it is the running process that is conscious, not the program itself as a static collection of instructions. This holds most clearly if we think of the program as an abstract mathematical structure, but it also holds even if we think of a physical realization of the program as a series of marks on a piece of paper or the corresponding contents of a set of computer memory cells. And even when a running of the program exists it is the *running process* that is conscious, not the program.

In order for the running to be conscious, it must be based on the sort of physical, reflexively-acting running argued for above. Notice here that it is not enough for a program to be able (when running) to examine or reason about its own computational states, or (at another level of description) its own beliefs, goals, intentions, etc. Existing AI systems can do this and are not thereby conscious (I strongly presume).

Rather, as a special case of Step 2 in section 2.2, the *physical runningness* of the program must "matter" to the program itself. Somehow, the progress of the computation has to be affected by the physical reflexively-acting as above. This could take various forms, but ultimately some state transitions in the program run must detect aspects of the physical running (causation) that physically binds some states together, and I would conjecture also that in some sense the binding can be affected by actions resulting from program instructions. Such detecting and affecting must arise through special physical properties of the device on which the program is being run, just as programs can interact with clocks, heat-sensors and so forth. However, it is not yet clear how intimate and complex the tangle of the reflexively-acting running and the computational transitions needs to be.

But, whatever the details, it follows that the full course of the computation is not determined by the program itself as a textual object, or binary-string object in computer memory, because the effect of that special physical running is not described by the program itself. It's a special, bottom-up physical effect that interferes with the program-ordained computational progress. Bottom-up physical effects in general are of course not foreign to everyday computation. For instance, hardware faults can affect a program run; or, more relevantly, the program might access the time values delivered by a clock in the system that measures physical time or temperature values delivered by a heat sensor in the computer. And such effects are not different in principle from non-program-ordained inputs coming from an external environment. It's just that the bottom-up effects I claim are necessary for consciousness are of a very special sort, in virtue of involving special reflexively-acting running.

The above may look like an argument that computation is not sufficient for consciousness. However, this is only so given a very abstract notion of computation that places few if any constraints on types of allowed implementation. Whether and how to constrain implementations is at the core of recent discussions about whether computation is observer-relative or not (Putnam 1988; Searle 1980, 1990; and papers in *Procs. 7th AISB Symposium on Computing and Philosophy: Is computation observer-relative?*, held at AISB50, Goldsmiths College, London, April 2014). Schweizer (2014), who represents one camp, cogently argues that computation in its pure mathematical definition is observer-relative, and is therefore not sufficient for accounting for mind, but it is possible to define a more empirically grounded computational theory of mind by implementing abstract computation in suitable physical ways.

As far as the Chinese Room is concerned, we should expect the overall system of the room and its contents not to be conscious, because unless special measures are taken, the implementation of the Chinese-understanding program is not implemented on a substrate of reflexively-acting running. Certainly, if the person or persons inside the room doing the individual computation steps are doing it consciously, then there is reflexively-acting running inside them. But this is insulated from the computation outside the people, in the room itself. It has to be that the reflexively-acting running can affect the outcome of computational steps. But if, by assumption, the people in the room are reliably executing the symbol manipulation rules, they are not letting the possible effects arising from their own internal consciousness affect operations outside themselves. In sum, the Chinese Room argument shows, not that there is no viable computational theory of (conscious) mind, but that normally considered ways of implementing a program or emulating its behaviour will not deliver consciousness.

3 Ways in Which the Proposal Helps Us

3.1 Help 1: (Multiple) Realization

A tenet presented in section 2.3, that no particular sort of physical stuff (such as biological stuff) is needed for consciousness, is a statement of multiple physical realizability of consciousness. But it is not merely a postulate, and appears to follow naturally from the idea of qualia arising from special, reflexively-acting runningness at a very low physical level. In principle, this may restrict the type of physical stuff that can be involved, but it is reasonable to think that the special runningness exists at much too fundamental a level of physical reality to affect the issue of whether, for instance, organic matter in complex brains is needed for consciousness.

But just as a claim that consciousness has to rest, say, in biological material is getting at the wrong level of physics, so talk of consciousness being able to result just from computation as traditionally described misses the crucial point that there must be *some* physical realization that has the special sort of runningness. In principle, it could be that the computational or dynamic organization of the substrate needs to be of a particular form, e.g., organized by a global workspace (Baars 1988; Dehaene & Naccache 2000) or as in Dennett's (1991) multiple-narratives account. But it has never been clear why such organization, no matter how complex, could not be present in a purely unconscious system. Steps 1 to 3 in section 2.2 argue that the system would not be conscious unless based on reflexively-acting physical running.

3.2 Help 2: What Has Higher-Order Thought got to do with Qualia?

The consciousness literature has been much concerned, on the one hand, with basic qualia such as pain and colour sensation and, on the other hand, with higher-order thought—believings, intendings etc, about one's own believings, intendings, etc. But do these two matters have anything to do with each other, in principle? Couldn't one exist without the other, especially basic qualia without higher-order thought? Wouldn't this in particular be what consciousness in some non-human animals must be like, if it exists? These questions are relevant to two related but distinguishable debates about the relationship between thought and qualia.

The first debate exists because of theories that qualia arise from sufficiently deep higher-order thought (Carruthers 2011; Van Gulick 2014). Such claims seem not to rely on any consideration of a particular sort of physical substrate for thought, but just on higher-ordedness itself. While I cannot do justice here to the extensive discussions about this idea and other aspects of Higher-Order Thought/Perception (HOT/HOP) approaches to consciousness, some remarks are in order. There are reasons to be sceptical that higher-ordedness of thought is enough generate qualia (although it may *help* to do so in some cases). An AI system reasoning about its own beliefs (via a simple modal logic of beliefs, say) could be reasoning about its beliefs about its beliefs, etc., to any degree of nesting, but it is doubtful that even many strong-AI advocates would claim that the system is thereby phenomenally conscious. Also, even if higher-order thought somehow leads to qualia intrinsic to thinking, it seems remarkable to propose that it is responsible for basic qualia such as pain, as this would force the strong step of excluding the possibility that creatures that have no higher-order thought can experience pain (unless there were an entirely separate way for pain to arise).

This paper's proposal provides an alternative view. It claims that reflexive action is central to the claimed special type of runningness, which is central to qualia. The fundamental form of reflexive action that this

paper proposes that runningness can have can be combined with suitably complex information processing (which would otherwise be unconscious) to deliver conscious, qualia-imbued higher-order thought. But the fundamental quale from the reflexively-acting running also supports less-cognitive or non-cognitive feelings such as pain, through complexes of proprioception, etc. So this is why both refined aspects of consciousness such as conscious higher-order thinking and primitive qualia such as pain both exist in human consciousness. They have a common cause.

And note that the reflexivity of the special runningness is so basic and non-cognitive that a consciousness may well not possess introspective thoughts, or thoughts at all, in the sense of involving relatively complex propositions. Thus the proposal allows living beings to have pain and vision qualia, say, even though the beings cannot think about anything, in the normal everyday sense of "thinking."

The second debate is about the existence of a special, purely cognitive phenomenology (see Bayne & Montague 2011). Some researchers (e.g., Tye & Wright 2011) claim that when there are qualia in thinking (as opposed to perceiving and emoting) they are actually the qualia that are found in perceiving (seeing, perceiving one's body state, etc.), engaging in imagery (visual, auditory, kinaesthetic, etc.), or emoting. Opponents claim that conscious thinking can have additional, "proprietary" qualia, distinctly different from qualia of perception, imagery or emotion. For instance, Shields (2011) claims that there are proprietary qualia involved in curiosity, wondering, remembering and so forth, while Robinson (2011) claims that there are only more "frugal" proprietary qualia such as certain forms of appropriateness, confidence, and affirming. The present proposal currently has less to say on this debate than on the first. As developed in this paper, the proposal neither supports nor undermines proprietary qualia. The only clear point at present is that, if the fundamental quale has a temporal quality (see section 4), then we get the prediction that qualia in conscious thought will have a temporal aspect, as do all other qualia (see again section 4).

3.3 Help 3: The Felt Difference and Generic Sameness of Qualia

Different qualia in human consciousness are of course different feelings—being in pain feels very different from seeing something. But there isn't a separate hard problem of consciousness for each one, possibly requiring distinctly different principles. It seems to be generally held that, if we could explain how, for instance, visual qualia arise in a physical universe, an essentially similar explanation, differing only in subsidiary detail, would apply to how other qualia arise.

The non-separation of the problem of qualia into separate, disconnected problems per quale no doubt reflects our commonsense intuitions that all our feelings have something in common—there is an intuition of a generic category of feeling. It seems reasonable to suppose that there is an intuitive hierarchy here. The quale of seeing red is different from the quale of seeing green, but both are special cases of a more generic quale of seeing-a-colour. Similarly, the different qualia of hearing are special cases of a more generic hearing-a-sound; and generic seeing-a-colour and generic hearing-a-sound are special cases of a generic feeling covering all qualia, which we could call feeling-something. But the question is, *why* is there such a generic feeling-something?

This paper's proposal points towards an answer to this question. Namely, all qualia somehow rest, at least in part, on the fundamental quale that is an aspect of physical reflexively-acting running. In the Introduction it was suggested that this quale may amount just to a feeling of existence, with an inherent reflexive quality of feeling-of-being-a-feeling. By way of an alternative, Step 4 in section 2.2 suggested that the quale could be described as something like "feeling something or other" where there is no sense of what sort of feeling

it is. The generic feeling-something of human consciousness could therefore be nothing more than the fundamental quale. Or, it might be an enriched, higher-level version of this quale, affected by concepts such as the self arising from higher-level cognition.

4 Temporal Consciousness and a Sub-Proposal

There is a natural suggestion one can make about the fundamental quale arising in reflexively-acting physical running. The very idea of "running" (the time-evolution of some portion of nature as governed by physical laws) suggests that the quale includes a sense of a moving now: a combination of a feeling of a now and a feeling of time passing. This is not imbued with any conceptualization of time or any sense that time is passing *for* any specific entity. Rather, it is the quale that might be described as that of sheer continuation. So, whereas without this step the quale could be called something like the feeling-of-something (cf. section 3.3), now we enrich this by adding a temporal aspect. The suggestion boils down intuitively to the idea that the fundamental quale consists just of a feeling of itself progressing through time.

According to this paper's proposal, all qualia are ultimately based, in a way yet to be clarified, on the fundamental quale. If this quale is temporal, then, in studying non-fundamental qualia, it is natural first to consider temporal qualia at the level of human common sense, such as everyday feelings that one is moving through time, or that time is passing slowly for oneself, and qualia concerning changing states such as movements in the environment or of one's own body. Our actual, everyday qualia of "now," "time passing," etc. derived in part from the basic, special runningness through many layers and tangled systems of proprioception and cognition about oneself and the world. This allows psychological variability, distortion and illusion in temporal qualia, e.g. being wrong about order or simultaneity, or having a varying sense of how fast time is passing.

A piece of supporting evidence for a central role for temporal qualia may be as follows. A person can lack qualia of pain, colour (or vision as a whole), smell, and so forth. Such people presumably normally still retain temporal qualia. The question is whether a person could lack all temporal qualia but still be conscious and thus have non-temporal qualia. The literature on consciousness suggests that there is general agreement that qualia of pain, colour, etc. inherently involve a sense of the feeling existing over a (possibly very short) period of time. The sense of something existing, even without change, over a period of time, such as an unchanging musical note or an unchanging colour, is itself a partly-temporal quale. Dainton's (2014) survey of temporal-consciousness research bears this out, despite the range of different theories about the nature of our temporal consciousness. Even theories that suppose that there is a series of instantaneous acts of awareness behind our temporal consciousness agree that our phenomenal experience always inherently includes impressions of duration and succession.

So, there seems to be something especially important about temporal qualia. This is immediately explained if the fundamental quale in this paper's proposal is temporal. It is not clear whether other proposals about consciousness can so readily explain why, e.g., a colour quale should only ever be experienced along with a sense of time passing.

Of course, one can be concentrating on things other than time, and be very unaware of how much time has passed or is passing. This perhaps arises in states described as "being in the flow." But is not clear that such states lack any sense at all of time passing: it's more that there's a feeling of time passing pleasurably and effortlessly, a feeling of being caught up in something with its own momentum. Indeed, it is revealing

that the metaphorical label used is "being in the flow" rather than "being out of the flow." Another famous type of case to consider is certain types of mystical state where there may be a sense of being above time and comprehending eternity. But it's not clear that the mystics in question are not still experiencing time passing: even if they perceive the normal earthly time dimension as a timeless unity, it is conceivable that they still experience a separate passage of time in their own consciousness. Because of the uncertainty about whether conscious experience without any temporal qualia is possible, the suggestion that the fundamental quale is temporal must remain more provisional than the main proposal.

It is nevertheless interesting to locate the suggestion in relation to the survey by Dainton (2014) of stances on temporal consciousness. Dainton categorizes the stances into Cinematic, Retentional and Extensional theories, each of which have multiple variants, and all of which have problems. There is no space here to discuss the range of possibilities, but in brief the present paper's proposal (including the present section's temporal sub-proposal) seems to fit best with the Extensional view, while also filling a possible gap in it. The Extensional view analyses temporal consciousness into intervals that take time and in each of which a succession of events in the situation being imagined or perceived (e.g., a ball bouncing) are co-represented in some way. It seems that phenomenology of duration, succession and change are just assumed to be inherent to these intervals, which have internal phenomenal unity. However, there is no explanation of this phenomenology. (This may sound mysterious, but the other types of account have their own mysterious elements.) This paper's proposal may help to found such an explanation, in that the phenomenal unity within an experiential interval could be based on the fact that running of any sort takes time, and the fundamental quale is *inherently* a matter of an instance of running that is reflexively acting upon itself over a period of time. Different phases within this instance could be bound up with representations of different moments in the situation being thought about.

5 Final Remarks

I have no *argument* that a suitable form of reflexively-acting running is sufficient for the presence of qualia. It is just a postulate. However, I believe that this postulate is more reasonable than, say, the idea that sufficiently complex, suitably organized abstract computation, when implemented in physical matter in familiar ways, can *ipso facto* be conscious. This comes down to the points in section 2.5, about the difficulty or impossibility of capturing, in a normally-implemented computation, the required type of reflexively-acting physical running (or, in more intuitive language, capturing the required reflexive form of physical causation). Rather, some special form of implementation is needed that involves the special reflexive running and that allows the computation to interact with it.

Notice that the argument for that need in Steps 1 to 3 in section 2.2 does not itself include the postulate that the reflexively-acting running constitutes a quale and is therefore *sufficient* for at least a primitive form of consciousness. That sufficiency postulate only comes in Step 4. The argument in Steps 1 to 3 merely seeks to show that the reflexively-acting running is necessary for consciousness. Thus, any account that does not feature such running, including any computational account where the implementational substrate doesn't feature it, fails a *necessary* condition whose support does not depend on the postulate in Step 4.

Acknowledgments

I am grateful to the reviewers for productive commentary. I am indebted also to Yasemin Erden for creating this special issue of the journal, and the workshop from which it was derived, giving me the impetus to express the ideas in this paper.

Bibliography

Baars, B. (1988) *A Cognitive Theory of Consciousness*, Cambridge: Cambridge University Press.

Baer, W. (2010) Introduction to the physics of consciousness, *J. Consciousness Studies*, **17** (3–4), pp. 165–191.

Bayne, T. & Montague, M. (eds) (2011) *Cognitive Phenomenology,* Oxford: Oxford University Press.

Bishop, J.M. (2009). A Cognitive Computation Fallacy? Cognition, Computations and Panpsychism, *Cognitive Computation*, **1,** pp. 221–233.

Carruthers, P. (2011) Higher-order theories of consciousness, in Zalta, E.N. (ed.) *The Stanford Encyclopedia of Philosophy* (Fall 2011 ed.), [Online], http://plato.stanford.edu/archives/fall2011/entries/consciousness-higher/ [1 Mar 2014].

Dainton, B. (2014) Temporal consciousness, in Zalta, E.N. (ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2014 ed.), [Online], http://plato.stanford.edu/archives/spr2014/entries/consciousness-temporal/, [7 Apr 2014].

Dehaene, S. & Naccache, L. (2000) Towards a cognitive neuroscience of consciousness: basic evidence and a workspace framework, *Cognition,* **79**, pp. 1–37.

Dennett, D.C. (1991) *Consciousness Explained*, London: Penguin.

Fekete, T. & Edelman, S. (2011) Towards a computational theory of experience, *Consciousness and Cognition*, **20**, pp. 807–827.

Husserl, E. (1991) *On the Phenomenology of the Consciousness of Internal Time (1893–1917)* (edited and translated by J.B. Brough), Dordrecht: Kluwer.

Leech Anderson, D. (2012) Causality-dependent consciousness and consciousness-dependent causality. *J. Consciousness Studies,* **19** (5–6), pp. 12–39.

Lloyd, D. (2012) Neural correlates of temporality: default mode variability and temporal awareness, *Consciousness and Cognition*, **21** (2), pp. 695–708.

Nunn, C. (2010) 'Landscapes' of mentality, consciousness and time, *J. Consciousness Exploration & Research*, **1** (5), pp. 516–528.

Nunn, C. (2013) On taking monism seriously, *J. Consciousness Studies*, **20** (9–10), pp. 77–89.

Primas, H. (2007) Non-Boolean descriptions for mind-matter problems, *Mind and Matter*, **5** (1), pp. 7–44.

Putnam, H. (1988) *Representation and Reality*, Cambridge, MA: MIT Press.

Robinson, W.S. (2011). A frugal view of cognitive phenomenology, in Bayne, T. & Montague, M. (eds), *Cognitive Phenomenology*, Oxford: Oxford University Press.

Schweizer, P. (2014) Algorithms implemented in space and time, in *Procs. 7th AISB Symposium on Computing and Philosophy: Is Computation Observer-Relative?*, held at AISB50, Goldsmiths College, London, 1–4 April 2014, [Online], http://aisb50.org/the-7th-aisb-symposium-on-computing-and-philosophy-is-computation-observer-relative/, [9 Apr 2014].

Searle, J. (1980) Minds, brains and programs, *Behavioral and Brain Sciences*, **3**, pp. 417–424.

Searle, J. (1990) Is the brain a digital computer?, *Procs. American Philosophical Association*, **64**, pp. 21–37.

Seibt, J.(2013) Process philosophy, in Zalta, E.N. (ed.) *The Stanford Encyclopedia of Philosophy* (Fall 2013 ed.), [Online], http://plato.stanford.edu/archives/fall2013/entries/process-philosophy/ [1 Mar 2014].

Shields, C. (2011) On behalf of cognitive *qualia*, in Bayne, T. & Montague, M. (eds), *Cognitive Phenomenology*, Oxford: Oxford University Press.

Skrbina, D. (ed.) (2009) *Mind that Abides: Panpsychism in the New Millennium*, Amsterdam: John Benjamins.

Talmy, L. (1988) Force dynamics in language and cognition, *Cognitive Science*, **12**, pp. 49–100.

Tye, M. & Wright, B. (2011) Is there a phenomenology of thought?, in Bayne, T. & Montague, M. (eds), *Cognitive Phenomenology*, Oxford: Oxford University Press.

Van Gulick, R. (2014) Consciousness, In Zalta, E.N. (ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2014 ed.) [Online], http://plato.stanford.edu/archives/spr2014/entries/consciousness/, [1 Mar 2014].

von Foerster, H. (1976) Objects: tokens for eigen-behaviors, *Cybernetics Forum: The Publication of the Amer. Soc. for Cybernetics*, **VIII**, 3/4, pp. 91–96.

Weber, M. & Weekes, A. (eds.) (2010) *Process Approaches to Consciousness in Psychology, Neuroscience, and Philosophy of Mind*, Albany, NY: State University of New York Press.

Weekes, A. (2012) The mind-body problem and Whitehead's non-reductive monism, *J. Consciousness Studies*, **19** (9–10), pp. 40–66.

Whitehead, A.N. (1929/1978) *Process and Reality* (corrected ed.), New York: The Free Press.

Whitehead, A.N. (1984) Time, in Ford, L. (ed.) *The Emergence of Whitehead's Metaphysics: 1925–1929*, Albany, NY: State University of New York Press.