Stepping down from mere appearance: Modelling the ‘actuality’ of time

Sally Shrapnel[*1], **Peter W. Evans**[†2], **and Gerard J. Milburn**[‡1]

[1] *School of Mathematics and Physics*, University of Queensland
[2] *School of Historical and Philosophical Inquiry*, University of Queensland

November 20, 2023

Abstract

In her paper, ‘The Open Universe: Totality, Self-reference and Time’, Ismael argues that the agent experience of dynamic time, in which the world comes into being, is more than ‘*mere* appearance’. The key to her argument is that the actual world includes agents (minds) and their corresponding agential (mental) activity, which Ismael describes as ‘the crucial step down from ‘mere appearance’ to actuality’. We present here a model of an agent as a complex physical system with a view to providing a tenable illustration of Ismael’s schema for understanding the ‘actuality’ of time. The radical implication for physical time as a result is that there is, in Ismael’s words, ‘no pure…conception of the way [the] world is independently of ourselves and our representational activity’, but rather only ‘embodied and engaged’ knowledge of the world.

1 Introduction

In the final passage of her paper ‘The Open Universe: Totality, Self-reference and Time’ (Ismael 2023: 16, emphasis in original), Jenann Ismael argues against the notion of a mind-independent world that bears ‘the way things *really are*’, one that shows that the agent experience of the world as coming into being ‘is *mere* appearance’. According to Ismael, ‘there is no well-defined conception of the Universe as a whole that is (for you) there *anyway*, *already*, or *independent* of your actions’, and the reason for this is that ‘the world *includes* minds and the effects of mental activity’. As Ismael states things, this move ‘is the crucial step down from ‘mere appearance’ to actuality’.

---

[*] email: s.shrapnel@uq.edu.au
[†] email: p.evans@uq.edu.au
[‡] email: g.milburn@uq.edu.au

We take the import of Ismael's argument to be that the agent experience of the world coming into being over time is an ineradicable, and indeed integral, ingredient in the actuality of time. In this work we take this approach seriously and consider a plausible representation of time according to a toy model of an agent as an embodied learning machine. We model an agent as a complex physical system that is constrained by the laws of thermodynamics, drawing energy from the environment to drive its actions and dissipating energy back into the environment as it *acts upon* and *detects* the world. Moreover, we take this agent to be able to *learn* about and construct its own representation of those dynamical processes in its environment that it can act upon and detect, which ultimately drives more energy efficient interactions between it and the the world.

To see more clearly the connection here to Ismael's argument, consider a more precise explication of this model of an agent. This agent is an open and dissipative physical system driven far from thermal equilibrium by an external low entropy source. It consists of a set of actuators to act on the world, sensors to respond to the world, and a learning machine to learn the functional relations between the physical states of the actuator and sensor. According to this model, learning is an entropy reducing process that alters the internal physical settings of the learning machine, based on feedback, to minimise the energy dissipation of the agent. The learned functional relations are encoded in the physical state of the learning machine, and constitute a causal model of the correlations between actuator and sensor records. This causal model is a representation of the 'laws of physics' in the agent's world, centred on the agent and its ability to act in its environment.

The significance of our toy model of a learning agent consist in the representation of time we take the agent to develop as a result of the learning process. An inherent part of establishing the correlations between actuator and sensor records during the learning process is that each sensor record becomes indexed with a corresponding actuator record. We argue that this indexing process is fundamentally a temporal ordering of the sensory data: the actuator in a sense functions as a relational clock for signals arriving from outside the agent. Since this ordering consequently underpins the learned laws, we claim that this process amounts to an implicit representation of *physical time* for the agent in terms of irreversible, dissipative internal dynamics – the agent represents the world as coming into being, entirely as a function of its specialised interaction with the world constrained by the laws of thermodynamics. Moreover, and intimately connected to Ismael's own schema for understanding an agent's representation of its environment, this representation emerges purely from the agent learning about its world with itself at the centre, driven by thermodynamic efficiency, and for the purposes of generating effective strategies to act in its environment.

Consequently, we take our toy model of an embodied agent, and our argument for how it represents time, to be a tenable physical illustration of Ismael's schema for understanding the 'actuality' of time. We conclude with the radical implication for physical time—time as it features in the physical laws—that there is, in Ismael's (2023) words, 'no pure. . . conception of the way [the] world is independently of ourselves and our representational activity', but rather only 'embodied and engaged' knowledge of the world.

2 Ismael's schema and the philosophy of time

Ismael (2023) proposes a schema for understanding our experience of the openness of the future: that is, that feature of our experience of time 'according to which potentialities seem to be transformed into actualities' (Penrose 1979: 591). According to this schema, the openness of the future is a consequence of an embedded agent attempting to model its environment while simultaneously modelling its own actions on that environment. The former task is an attempt to represent a domain that is apparently external to, or independent of, the agent; the latter is an attempt to represent the agent itself and its actions in that domain. This inescapable self-reference the agent makes to itself in its modelling of its environment leads to what Ismael calls 'interference',[1] which, while not constraining the agent physically, 'does place constraints on what the system can truthfully represent and it does lead to an essential incompleteness in the worldview of an embedded Agent' (Ismael 2023: 9). This incompleteness of representation manifests, Ismael claims, as an open future for such agents.

Ismael's project can be seen as fitting within the debate in the philosophy of time concerning how we account for two very different sources of evidence about the nature of time. On the one hand, our direct experience of time provides apparent evidence that time is in some sense dynamic, that the present moment is in some sense privileged, and that the past is fixed while the future is open. We call this characterisation 'manifest time'.[2] On the other hand, the description of the world that we get from the physical sciences provides apparent evidence that time is in some sense static, that no moment of time is more privileged than any other, and that the past and future are equally real. We call this characterisation of time 'physical time'. In so far as the best guide to the nature of time is our best physical description of time, this evidence seems to provide a compelling argument against the veracity of manifest time. One way to understand Ismael's project is providing a basis for challenging this state of the debate: we can conceivably understand manifest time as arising from the physical structure of the world, so long as we include the embedded, embodied agent in that structure. To explore Ismael's schema in more depth, let us say a few more words about physical time.

The standard means in the physical sciences of formally describing the physical world is through dynamical laws: theories and models parameterising dynamical processes in nature. By definition, such 'dynamical' laws are 'in time': the laws are formulated in

---

[1] 'If the domain includes its own activity then some of what happens is stuff that it does; and the more connected the domain, the more difficult it will be to quarantine the effects. Interference impedes pure knowledge acquisition.' (Ismael 2023: 3)

[2] This terminology is employed by Callender (2017) in his analysis of the nature of time, and is derived from the 'manifest image' of Sellars (1963).

terms of a real parameter—physical time—conventionally denoted as t. It is common in both physics and the philosophy of time to regard t as time itself. However, when we consider the way that scientific agents come to these sorts of dynamical models, we see that the *measurement* of time plays a latent role in the agent's formulation of physical time. In so far as the measurement of temporal duration is achieved by clocks, we claim that the nature of physical time can be better understood as intimately connected to the nature of clocks (Evans et al. 2023).

Moreover, by taking seriously the measurement of time, we place ourselves squarely in the philosophical tradition that takes time to be relational. According to this tradition, t is taken to be a kind of surrogate parameter for a relation between changes in values of physical (typically spatial) variables (see, for instance, Mach (1911)). As we argue more fully elsewhere (Evans et al. 2023), given a rudimentary account of a relational clock as simply a dissipative, nonlinear oscillator with a cycle counter that coordinates coincidences between local events, it becomes clear to see that the process of parameterisation in the context of relational clocks cannot be interpreted (as is typical for parameterisation) as uncovering a hidden cause of the correlations between the coincident local events: 'time' does not *cause* the dynamics of either system between which local events are coincident, including the clock. Rather, parameterisation by physical time is merely a mathematical convenience, introducing a surrogate parameter to stand in for whichever local physical system is employed as a relational clock, and which can then be used to parameterise the behaviour of physical systems in the service of scientific modelling.

Given this insight into the connection between physical time and clocks, in the next section we outline a toy model of an agent that learns the 'laws of physics' in its environment. It does so by modelling the correlations between its actions in the world and its sensations of the world as a set of functional relations that allow it to emulate 'dynamical' behaviour in the environment. This process of learning embodies an internal physical clock through a relative ordering of action and sensation events. We argue that this amounts to an indexing of sensation events by action events, and this indexing is fundamentally a temporal parameterisation of the internal physical states of the agent. Since this parameterisation consequently underpins the learned laws, we claim that this amounts to an implicit representation of physical time for the agent.

3 An embodied agent modelling time

In a pair of recent papers (Evans et al. 2021; Milburn et al. 2023), we develop a toy model of an embodied agent as a complex physical system (Fig.1). This agent is an open and dissipative physical system driven far from thermal equilibrium by a temperature gradient. It consists of a set of actuators to act on the world, sensors to respond to the world, and a learning machine to learn the functional relations between the physical states of the actuators and sensors. The temperature gradient ensures these specialised subsystems are

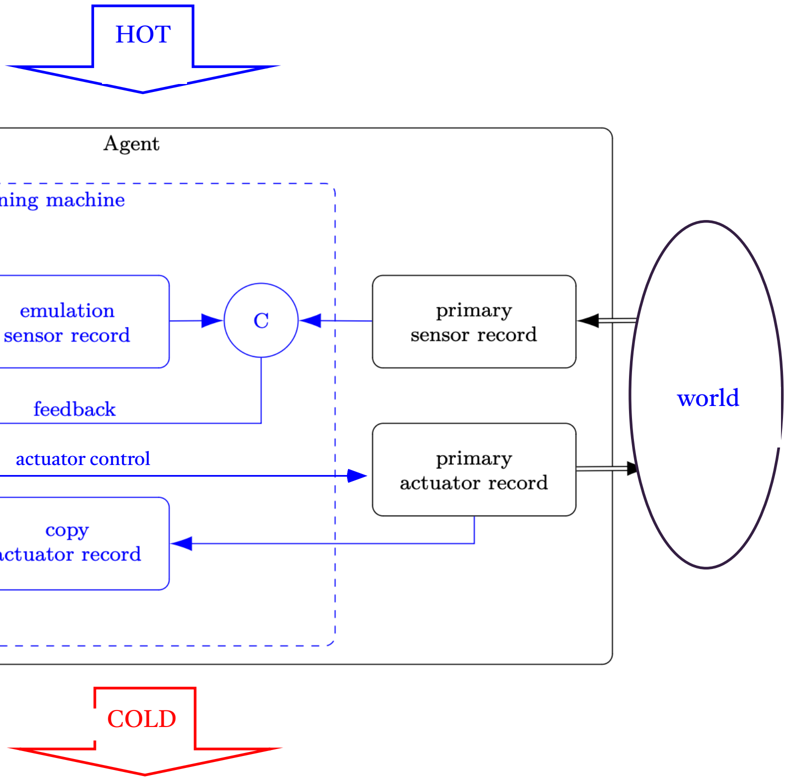

Figure 1: Schematic diagram of the learning agent. A temperature gradient drives the agent away from thermal equilibrium, dissipating energy and powering sensors, actuator and the learning emulator. The actuators do work on the world while the world does work on the sensors. The learning emulator dissipates energy to reduce its entropy and minimise the overall energy dissipation of the agent by minimising its prediction errors.

in a non-thermal-equilibrium steady state. During an interaction between the agent and the world, actions originate in the agent by a change in the physical state of an actuator, which is then able to do work (in the thermodynamic sense) on the environment and, in parallel, do work on the internal learning machine (Fig.1), whose task it is to simulate an expected sensation. The simulated sensation is then compared to the actual sensation coincident on a corresponding sensor. If they are not (approximately) the same, the agent changes the internal state of the learning machine before the next action/simulation event. As such, learning is an entropy reducing process that alters the internal physical settings of the learning machine, based on feedback, to minimise the energy dissipation of the agent. An appropriately engineered learning machine will eventually thermodynamically drive simulated sensations to match, with little error, the actual sensations.

The role of the learning machine is thus to compress the *correlations* between the signals sent to the actuators and received from the sensors into functional relations, whatever form the signals may take. The learning machine thereby learns, through the agent's interaction with its environment, a series or network of functional relations between variables taken to represent specific types of actions and sensations. The learned functional relations are encoded in the physical state of the learning machine and, we claim (Evans et al.

2021; Milburn et al. 2023), can plausibly be interpreted as causal relations attributable to external phenomena, i.e. a *causal model*, in precisely the way that Bayesian networks can be given a causal interpretation (Woodward 2003; Pearl 2009). In so far as these correlations between actuator and sensor states tell the agent something about the external world, the physical state of the learning machine then embodies the causal relations between inputs to the world from the actuators and outputs from the world to the sensors. Significantly, the learned causal model will be inherently asymmetric, in the sense that it models only that dynamical structure of the environment that is exploitable by the agent for bringing about desired effects (as given by the model). In this context, the learned causal model represents the 'laws of physics' in the agent's world, centred on the agent and its ability to act in its environment.

For the learning machine to operate as intended, sensor records must be indexed with a corresponding actuator record, otherwise there is no guarantee that the agent will be able to match corresponding actuator and sensor records, and so no guarantee that the agent will 'discover' exploitable dynamical structure in the environment. We claim that this indexing process is fundamentally a temporal ordering of the sensory data: the actuator in a sense functions as a relational clock for signals arriving from outside the agent. The functional relations that the agent then learns that codifies the correlations, and so provides the 'laws' in the agent's world, implicitly parameterises the set of locally ordered coincidences between internal physical variables in the actuators and sensors. Since this consequently underpins the learned laws, it is natural to interpret this parameterisation as a representation of 'physical time' for the agent in terms of irreversible, dissipative internal dynamics: it is the time parameter t of the best physical description the agent can give of its environment. Significantly, though, this best physical description is derived from a set of interactions between agent and environment in which the world comes into being dynamically over time. Moreover, this representation emerges purely from the agent learning about its world with itself at the centre, driven by thermodynamic efficiency, and for the purposes of generating effective strategies to act in its environment.

It is worth noting briefly at this point that, since physical time for the agent here arises purely from the operation of the agent's specialised subsystems, there is no need for a global clock inside the agent that tracks (or otherwise) physical time 'out there' in the world. As such, an agent that produces actions at random times (as measured, say, by an external clock) would be just as effective in coordinating internal coincidences as an agent that incorporates an explicit global clock by which to operate. The kind of internal relational clock we describe here already has a recognised basis in neuroscience: 'It is hard to justify the hypothesis that there are neuronal networks whose sole function is to clock time' (Buzsáki 2019: 261).

4 Stepping down from mere appearance

We contend that our toy model of an embedded, embodied agent learning the laws of its world provides a neat illustration of Ismael's schema. The model furthermore neatly captures the aligned thermodynamic, causal, and temporal arrows of the world as seen from the perspective of an embodied agent. The key to the toy model is that the embodied agent is limited to learn about those parts of the world that are accessible to its specialised subsystems, and so learn about the world through its own actions. In this way, the causal model it learns inescapably has itself and its own actions in its local environment at the metaphorical centre of the model. It is worth dwelling on this point briefly.

Consider the embodied agent accumulating information about its environment through passively sensing the world; that is, the agent refrains from using its actuators. Provided that the agent can obtain data concerning which events might be coincident, and which events come before or after other events, it is conceivable (although we do not proclaim to show this here) that the agent could construct a Bayesian network over the variable types it detects and, under appropriate statistical conditions, interpret any concomitant functional relations as comprising autonomous causal mechanisms that provide causal information about possible actions it can undertake in the world (as per the structural account of causation (Spirtes et al. 2000; Pearl 2009)). However, as is well known in the philosophy of causation literature, information gained via interventions is a key ingredient for interpreting any particular functional relation from the network as a *causal* relation (Woodward 2003). As a consequence, for an embodied agent in a complex environment, it is exceedingly difficult for the agent to model the world, particularly for the purposes of exploiting the model to bring about desired effects in the local environment, independently of its own actions. We take this to be, then, a plausible exemplification of the inescapable self-reference with which embedded agents must contend in Ismael's schema when modelling themselves and their actions as part of the world: their model will invariably contain a signature of themselves.

More importantly, however, this provides a clear illustration of what Ismael (2023: 16) means by the claim that 'there is no well-defined conception of the Universe as a whole that is (for you) there *anyway*, *already*, or *independent* of your actions'. In so far as the learned model of the 'laws of physics' in the agent's world constitutes the agent's best physical description of its environment, this learned model inescapably contains the agent and its actions as the perspective from which 'physics' describes the world. So when such an agent extracts from physical time, based on the parameterisation of its laws, 'the way things *really are*' and so takes physical time as a good guide to its metaphysical investigation into the nature of time, it might proclaim that, say, time is static, has no privileged moment, and the future is as equally determined as the past, and thus the experience of the world as coming into being is mere appearance. But given all that we have said about how to interpret the agent's own model of its world, we can step down here from mere appearance to actuality. While we do not intend this toy model to necessarily capture what it means to be an embedded, embodied, learning *human* agent, if we were to take ourselves to be just such an agent, then the natural conclusion of our argument is best expressed by Ismael's (2023) own words: 'There can be no pure and complete conception of the way that world is independently of ourselves and our representational activity', but rather only 'embodied and engaged' knowledge of the world.

References

Buzsáki, György (2019) *The Brain from Inside-Out.* Oxford University Press.

Callender, Craig (2017) *What Makes Time Special?.* Oxford University Press.

Evans, Peter W., Gerard J. Milburn, and Sally Shrapnel (2021) 'Causal Asymmetry from the Perspective of a Causal Agent', philsci-archive.pitt.edu/18844/.

Evans, Peter W., Gerard J. Milburn, and Sally Shrapnel (2023) 'How Clocks Define Physical Time', philsci-archive.pitt.edu/22565/.

Ismael, Jenann (2023) 'The Open Universe: Totality, Self-reference and Time', *Australasian Philosophical Review*: 1–16. doi:10.1080/24740500.2022.2155200.

Mach, Ernst (1911) *The History and the Root of the Principle of Conservation of Energy*, Philip E. B. Jourdain, trans. The Open Court Publishing Co.

Milburn, Gerard J., Sally Shrapnel, and Peter W. Evans (2023) 'Physical Grounds for Causal Perspectivalism', *Entropy* **25**, 1190. doi:10.3390/e25081190.

Pearl, Judea (2009) *Causality: Models, Reasoning, and Inference*, *2*nd edition. Cambridge University Press.

Penrose, Roger, (1979) 'Singularities and Time-Asymmetry', in Stephen W. Hawking and Werner Israel, eds., *General Relativity: An Einstein Centenary Survey*: 581–638. Cambridge University Press.

Sellars, Wilfred (1963) *Science, Perception and Reality.* Routledge & Kegan Paul.

Spirtes, Peter, Clark N. Glymour, and Richard Scheines (2000) *Causation, Prediction, and Search.* The MIT Press, 2nd edn.

Woodward, James (2003) *Making Things Happen: A Theory of Causal Explanation.* Oxford University Press.