

Author Response to Commentaries: 'The Open Universe: Totality, Self-reference, and Time'

Jenann Ismael

To cite this article: Jenann Ismael (2024) Author Response to Commentaries: 'The Open Universe: Totality, Self-reference, and Time', Australasian Philosophical Review, 8:3, 291-313, DOI: [10.1080/24740500.2024.2522398](https://doi.org/10.1080/24740500.2024.2522398)

To link to this article: <https://doi.org/10.1080/24740500.2024.2522398>



Published online: 28 Aug 2025.



Submit your article to this journal [↗](#)



Article views: 13



View related articles [↗](#)



View Crossmark data [↗](#)



Citing articles: 1 View citing articles [↗](#)

REPLY



Author Response to Commentaries: ‘The Open Universe: Totality, Self-reference, and Time’

Jenann Ismael

Johns Hopkins University

I’m so grateful to Heather Dyke who edited this issue and to the commentators whose made the paper seem like a house with a lot of doors opening outward. I’ll make some general remarks that will be helpful to have up front before turning to the commentaries individually. There is little that I disagreed with and much that I learned from.

In the paper I introduced the concept of interference and argued that:

1. Any system representing the world from within will encounter interference.
2. The purest form of interference is self-reference.
3. Indirect forms of interference arise because of connections to the wider physical landscape.
4. Interference is different from ignorance and disturbance.
5. Interference has to be taken into account in our understanding of our place in the world.
6. Added to a classical world with a thermodynamic gradient, interference gives us a way of understanding a sense in which the future is open and can provide an interpretation of what Penrose calls ‘the arrow most difficult to comprehend’.

Interference is best characterized by examples and in contrast to ignorance and disturbance.

Examples of ignorance

1. My view is blocked by an obstacle.
2. Something lies outside my perceptual horizon.

In these cases, there is something there to be known, independently of what I believe or think about it, that is, independent of my representational activity. It happens that I’m just not placed to ascertain it.

Examples of disturbance

1. I want to know if the light in the fridge is on, so I open the door to check.
2. I check the pressure in a bicycle tire but allow air to escape as I connect the gauge.

In these cases, there is some fact there to be known (the state of the system before the interaction), but the result of the measurement doesn't reflect that fact.

Examples of interference

1. I ask the computer: Is the answer that is about to appear in the output channel 'no'?
2. Election outcome predictions that affect how people vote.
3. Counter predictive algorithms take a prediction of their output as input and output the opposite.

In cases of interference, the fact being represented cannot be stabilized independently of the representation (or representational act). Self-reference captures the logical structure of interference in its purest and most direct form. There's no getting around that if your representational activity falls under its own scope then part of what you are representing is your representational activity itself, and there will be at least some things that can't be stabilized independently of how they are represented. It arises in less direct forms because our representational activity is connected to other events; we can't disentangle our representational activity from the wider world. On a personal level it is connected to what we do and say, and that is connected to what happens in the world and to what other people believe and do. From a broader perspective, look around any human dwelling or city, the physics of the environment (with its refrigerators and roads, its buildings and cars) cannot be understood without understanding the role of representational activity in shaping the material environments in which it is contained.

The phenomenon of interference is undeniable. I'm exploring interference in a physical setting, but most philosophers will have encountered it first in the formal setting where it is the source of the paradoxes for formal logic. Consider a space of propositions connected by inferential relationships. Add a representation relation that ranges over the space itself so that some of the propositions refer to themselves. Introduce a truth predicate and we have the makings of negative interference in the form of a proposition that asserts its own falsity. The setting is different, but it has the same formal structure. When one is representing a domain from within and aiming for full coverage, interference is inevitable. It will show up explicitly if one attempts to arrive at a complete and accurate representation of the domain.¹ Whether it can be contained (kept from infecting other elements in the domain), and what its significance is, will all depend on the setting.

We can explore it in *different* physical settings (for example, in a classical regime or a quantum one). We can explore it in virtual worlds or in formal systems. We can explore it in board games or computer games. Chess provides an especially illuminating example because it reproduces many of the features of our world.²

The general claim of the paper is that interference is an unavoidable and basic aspect of our perspective on the world, and that our experience of the world is organized around it. The specific claim is then that if we look at how interference plays out

¹ There is a weak kind of completeness that is unproblematic: completeness and accuracy in positive respects. Negative interference arises if we include accuracy in negative respects, demanding truthful answers to all yes/no questions of fact.

² Rovelli (2024) invokes it in his commentary, and I'll say a little about it below.

against the background of the thermodynamic gradient, we arrive at an understanding of the sense in which the future is open. I'm using interference, that is to say, to give us an understanding of Penrose's 'most difficult arrow'.³

The fact that we are part of the world and that our representational activity is woven into the fabric of the world that we are representing is what gives rise to interference. I am looking at how it manifests in a dynamic setting in which the basic elements are events, assuming the classical dynamical laws and a low entropy past. Why these? because they structure the physical context against which our own lives play out, and so they regulate how interference propagates at the level that we engage with the world. They are part of the explanatory infrastructure of the time-directedness of the world as we find it.⁴

That is the heart of the paper. I started out wanting to understand the difference that we draw intuitively between time and space. Why is it that we think of space as a fixed object that comes into view in stages, but tend to think of time as coming into being as it is experienced? What is it about our relationship to time that underwrites that intuitive difference? What can we learn from it?

Nota bene that trying to interpret the inability to stabilize the future as *merely epistemic*—that is, trying to confine it to the appearances/keep it quarantined at the level of representation and think of the world itself as 'already there', 'there to be known', fully defined in all respects and merely coming into view—fails to appreciate the difference between ignorance and interference. It assumes a separation between the representation and reality that breaks down because our representations are part of the world. Our representational activity is woven into the fabric of the very reality that it represents in a way that makes a clean separation impossible. This is crucial to understanding the paper.

Price (2024), in his commentary, raises a thought experiment involving angels that observe the world but don't intervene, so that their representational activity is passive and unconnected to the events they represent. The imaginary example provides a helpful contrast with our own situation and helps us understand why—in our own case—we can't just exclude our representational activity from what we mean by 'the world'. It tells us why we can't cordon off our representational activity, confine interference to our mental lives. What you think affects how you behave and your behaviour affects the landscape around you; that means you can't simply lift your representational activity out of the world and leave the rest intact. This underscores the fact that your

³ I'm working here in a physics tradition that disassembles the world of common sense with its primitive causal arrow and arrow of Becoming and reassembles it from a physics whose basic elements are given by physical theory (in this case, a relativistic space-time, time symmetric laws, a thermodynamic gradient).

⁴ If we want to explain any very general feature of our experience, we should hold fixed things that are part of the unvarying background against which we live our lives. The Past Hypothesis alone isn't enough. A lot of additional implicit scaffolding is needed to generate a world with the kind of emergent causal structure that rewards information gathering and utilization. Weinberger, Williams, and Woodward (2023) attempt to characterize some of the additional scaffolding. Here's how I think about this: think of the fundamental laws as generic, all-purpose constraints on what can happen. In addition to laws, we need special bespoke constraints that create platforms for local action. Generically, if we want to understand what happens in our world over the time scale of billions of years, we add conditions like a thermodynamic gradient. If we want to know how things behave close to the surface of the earth, we add a gravitational field. We might also add local constraints that act as scaffolding in particular situations. What conditions we treat as constraints in modelling a situation will depend on what we want to model, and the time and spatial scales we want to have in view. Generally, we want to choose constraints that reflect what is fixed over the time and spatial scale we are interested in (and can be treated as part of the background, not part of the changing action).

representational activity isn't passive but plays an active role in shaping the future. Our thoughts and decisions are an essential and integral part of the law governed fabric of the world. The internal process that transforms a set of potential actions into a definite decision is *part of the worldly process that transforms potentiality into actuality*. And that means that the world really is playing out *in real time* as our thoughts and decisions unfold. The proper time of my experience—that is, the time in which my decisions unfold and in which I go from uncertainty, through indecision to resolve—is wound into the external world of matter in motion. It is one and the same as the time of the external reality I perceive and represent.⁵ The output of my decisions is fed into the world and then back to me through perceptual channels when I see the effects of my own movements. That is what gives meaning to the sense of openness, of urgency, of not-yet-fixedness of the future. That is what it means for the *world* itself (and not just my experience of it) to be in process, not yet achieved, still unfolding, not there to be known, but swinging in the wind between different possible futures as I waver between doing this and doing that. Or so I suggest.

So the way to think of the progression of the argument here is that we start with an interpretation of Becoming in our experience and transfer it to the world by weaving our behaviour into the physical development of the world in a way that keeps us from quarantining it as a merely epistemic effect (that is, merely a feature of the appearances.)

Another upshot of taking interference seriously is the unavailability of a god's eye view. The Universe—understood specifically as the totality of what happens—cannot be (for us) well-defined in all respects. For us, the Universe has an irremediable sort of incompleteness. It is not an accident that what we do to stabilize the world as an object of knowledge often involves imaginatively casting ourselves in the role of an observer from the outside, detaching it from our input.

For most scientific purposes, none of this makes a bit of difference and it is fine to suppress interference. It makes perfect practical sense to adopt a god's eye view in cosmology, for example, that is, to pretend that we are looking at the Universe from the outside and it is well defined in every respect independently of our representational activity. Interference is not going to make any difference to the large-scale structure of the universe, Einstein's field equations, or almost anything cosmology cares about. It often matters is philosophy, however, where the god's eye view is invoked to correct some feature of our experience: for example, to tell us that the passage of time, or the openness of the future, or Becoming (properly understood) is illusory. I am not saying that time itself flows, or that the Universe itself is undergoing an absolute and irreversible process of coming into being, or any of the other ways of reifying Becoming at that level of the universe. I am saying rather that there is a very real sense in which our own futures are open for us, that the Universe is (as of now, and indeed until the end of time) incomplete, and the world is coming into being as we experience it. And I object to the tendency to give the god's eye view priority and treat it as revealing the way things *really* are. The god's eye view is not autonomous. It is partly constituted by the representation-guided activity of creatures like you and me and can't be detached from it while remaining complete. There's nothing illusory about that.

⁵ I'm taking on board here the lesson of general relativity that fundamentally there is no time but proper time. The point I'm making in the text is about bridging internal time and external time.

It is worth noting that how interference manifests itself is a perspectival phenomenon. How one person's beliefs and deliberations feed into the world (what we might call their 'practical perspective' on the world) is different from another's. It is also separable from their perceptual perspective. Consider, for example, two chess agents (human or virtual) pitted against one another. Chess is a game of perfect information. Both players have full information about the configuration of the board at all times and that information is presented in a non-perspectival way that simply list the positions of pieces. Even so, they can (and should) have different probabilities for future events and update on the same information in different ways. Black will take an epistemic attitude towards the movements of white pieces (it will use information about past regularities and so on to predict those movements and will correct bad guesses as information comes in). It will have no uncertainty about the movements of black pieces, however, and will use its own deliberations strategically to guide the game in the direction of a Black victory. Same, *mutatis mutandis*, for White. *The difference here is irreducibly one of how their own deliberations feed into and interfere with the unfolding events.*

Turning now to the commentaries. Here's the TLDR to orient you:

- There were two commentaries that absorbed the central point and deepened it: Rovelli and Deng.
- There were two expansive commentaries that applied the ideas of the paper to research programs that take seriously the idea that information gathering and utilizing systems are part of the physical universe. Shrapnel, Evans and Millburn, in their commentary, show how the schema for information-gathering and utilization is physically realized and why the internal arrow of a system engaged in it aligns with the thermodynamic arrow. Adlam, in her commentary, argues that applying interference not just to individual agents, but to networks of communicating agents provides a global temporal ordering, thereby solving the difficulty of extracting a global ordering for two research programs in modern theoretical physics (the matrix process formalism and quantum reference frames).
- The remaining commentaries challenge various elements of the paper.

Price (2024) argues that my characterization of the agential perspective is too thin. He thinks we don't have agency until we have deliberation and that the internal arrow of an agent that deliberates about what to do is what explains our inability to affect the past because 'what we really mean by "the future"' is the direction towards which we deliberate.

Hoerl (2024) interpreted me as arguing that the flow of time is integral to what it is to have a conscious point of view and discusses that position.

Braddon-Mitchell and Miller (2024) argue that the point about relativity near the end of the paper wasn't enough to establish human freedom, and that interference wasn't by itself enough to explain temporal experience.

Zhu (2024) argues that the discussion relies on an unjustified assumption that slips from instability of appearance to instability of fact.

I agree with some of this but not all; details to follow.

Price's work on temporal asymmetries, causation and the 'ultimate contingency' (a phrase he borrows from Frank Ramsey that licenses treatment of one's own actions as a free variable) introduced me to these topics and is at the root of these

ideas (Price 1992). I've known Huw since 2005 (I spent 5 years at the Center for Time, which he started and directed). This paper—like so much of the work I've done since that time—is a footnote to his work. If you aren't steeped in Price's work, steep yourself (Price 1996, 1998, 2011, 2013).

Deng

Deng's (2024) discussion captures—much better than I—the way that interference alters (by undermining) the A-series/B-series dichotomy of time. She's right that I'm simultaneously trying to give an *interpretation* of Becoming (trying to characterize the phenomenology) and say how it arises. She is also right that this piece builds on my other work on tenseless time and relativity-friendly passage. Instead of trying to paraphrase the commentary here, I want to encourage you to read it. I found it deeply illuminating myself.

There is one part of the discussion that I want to highlight in particular. Starting with the familiar contrast between two ways of thinking of time, codified in McTaggart's A-series and B-series. The A-series is the events of history ordered by past present and future, with the passage of time understood as the continuous movement of events from future through present to past. The B-series is the set of events ordered with respect to earlier and later. Deng notes rightly that I reject both the A-series and B-series as complete or autonomous conceptions of time and replace them with two nearby views that she denotes 'pre-B' and 'pre-A', respectively: *viz.*, the view of the Universe that is invariant under transformations between internal perspectives and the internal view of the Universe, which is comprised of a stream or sequence of views from particular temporal perspectives. The pre-B and pre-A views do not compete, but are rather two sides of the same coin, both rooted in our experience and appropriately deployed in different settings. When we are doing physics, the emphasis is on the abstract structure of invariant relations between moments in time. When we are living—processing sensory information, making decisions, actively engaged in trying to shape the future—it is the pre-A structure that is in the foreground. Neither is complete without the other, and they are inseparable in a particularly strong way.

She writes:

the self-reference view ... wants both to take seriously the original metaphysical A versus B opposition, and to transcend (or overcome) it. ... The question of whether such a thing really exists in temporal reality has occupied analytic metaphysicians for many decades. At the same time, it transcends that opposition, by transforming both the A- and the B-theory, in different ways. The B-theory turns out not to exist; it's a sort of intellectual mirage. And the A-theory is re-interpreted to fit neatly into pre-B, the view we take anyway when doing physics, stripped of metaphysical baggage. We are left having genuinely addressed a metaphysically rich explanandum, with truly physics-friendly resources, not just as an illusory experience, but as part of temporal reality. (Deng 2024: 276)

And I think that her suggesting that we have to go back to a pre-A-theoretic/B-theoretic structure, before the fissure was made, in order to reconcile them with one another is deeply insightful. I would extend the lesson by saying that it alters (by undermining) the separation of appearance and reality quite generally.

She ends the piece with a question:

How can A-intuitions be given this new content? Can we restate the psychological arrow as a ‘relentless sense of being part of Universe’ or similar? (Why ‘relentless’? How is it still an arrow?) On the other hand, without the meaning-giving element, how does self-reference accomplish more than locate Agents’ thoughts in the manifold while reminding us that they are ours?

I’m not sure that I can give a very sharp answer. Physics tends to make us reflective about aspects of our experience that we take for granted in everyday life. The pre-reflective sense that the future is open isn’t very clearly conceived; it is more amenable to images and metaphors than analysis. It clearly has something to do with the idea that the future can be affected by what we do in the here and now, whereas the past is beyond control. It becomes a puzzle in the context of a relativistic theory, where the basic reality is a four-dimensional manifold, to say what the difference amounts to and how it can be given literal content. It is in this context that we are forced to subject our experience to analytic attention and ask, ‘what does it mean, or what does it amount to, to say the past is fixed and the future is open?’. Our experience is ultimately our only source of information about ourselves and the world and the goal is to give us a way of understanding why things seem as they do, assuming that the world is as physics describes it, recognizing that ‘how things seem’ is a soft target of analysis, one that shifts under reflective attention, is pervaded with belief, cognitively penetrated and difficult to pin down. This is a kind of difficult hermeneutic process that falls between description and interpretation. There are aspects of our experience that can be dismissed as artefacts of mistaken belief; but there are some that seem too central and universal and resistant to correction not to be telling us something important. This I believe is one of those, but that is a partly interpretive conclusion. There is no simple or sharp way to know which is which.

Rovelli

Rovelli perfectly understood what I was trying to do, was right about all of the substantive points, and spot on about where to place the emphasis. This is a very rich piece that goes well beyond mine in ways that I want to elaborate. In discussing the example of a chess game, Rovelli comments:

For a strictly deterministic computer that plays chess and explores many alternative possible developments of the game before choosing its move, the future is, in a very literal sense, open: the meaning of this openness is precisely the existence (within the representation of the game that the computer holds and embodies) of the ensemble of possible future developments of the game itself, that the computer is exploring. Second—and this is the key point that Ismael emphasizes—the future is actually going to be partially determined exactly by the deliberations itself of the agent. Deliberation here is a physical process, it is a physical process that requires a description as considering alternative futures, precisely because deliberation itself contributes to determining the future. (Rovelli 2024: 261)

Precisely right. The assessment of these futures in the hypothetical mode is part of the process that determines the outcome so for the system carrying out this assessment, there is implicit self-reference built into any attempt to predict its own future. That implicit self-reference shows up: (i) by creating the potential for negative interference that prevents it from truthfully answering any yes/no question of fact, and (ii) by creating positive interference: it opens up multiple solutions in the future direction. What do I mean that it opens up multiple solutions in the future direction? Suppose a chess

program that controls white is trying to calculate how the game unfolds. If it is calculating the forward development of a game that it is not playing—even another chess agent running the same program and started in the same state—everything is smooth and unproblematic. If it is trying to calculate how the game that it is *playing* unfolds, when it tries to do the computation and gets to the stage of the computation that predicts its own first move, it will hit a conditional branch whose outcome depends on its own decision. At this point alethic constraints fall away. There are multiple consistent solutions. It can't calculate its first move without *predicting* its decision and there is no way for it to give a false answer. 'I decide to P' is self-fulfilling. It is performative. It makes itself true.⁶ All of this holds even in a deterministic setting.

Suppose I'm playing for white and you're playing for black. I have information about how I'm going to react conditionally. (I'm being generous here about what it means for me to have that information; if I can compute it from information I have then I'll say that I have that information. And I'm going to assume I have information about my own physical state and anything in my past.) To compute what I'm going to do at $t + 3$, however, since what I do will depend on what you do, I have to have information about what you do at $t + 2$ before I can make my decision. But since what you do at $t + 2$ will depend on what I do at $t + 1$, and I don't have that information until $t + 1$, I can't compute anything until I decide. That is true at every stage. It is because the inferences are all conditional, and conditional on what I decide, that the whole future of the game hangs in abeyance at the beginning and plays out in steps that I can't compress. I decide and wait, and then decide again and wait again. And the same, *mutatis mutandis*, for you. Each of us is at every turn doing the computation that will at once tell us what the future will be and determine the future. The thing that really keeps us from stabilizing the future of the game in advance—the thing that keeps us from getting ahead of ourselves—is that our predictions feed into our decisions and interfere with what we do. As long as predictions are making a difference to what happens at other times and places, interference will infect those other times and places.

Now let me connect this to the point that Rovelli makes about the evolutionary story and teleological thinking. He is right about this as well. The larger physical setting traces the existence of IGUS's (information-gathering and utilizing systems) to the special opportunities that the thermodynamic gradient presents to exploit the availability of information to guide action strategically. There are two sides to this; energy and information. A physical IGUS has to extract free energy from the environment to maintain its physical integrity. In the service of this, functionally, it is designed to use information to guide behaviour. The thermodynamic gradient is crucial to both. The two aspects have been studied in largely separate literatures.⁷ The illuminating thing about an IGUS is that it provides a schema of a physical system that embodies all of the arrows whose physical alignment we want to understand: the physical and the representational, energy and information.⁸ The features of our own mental lives

⁶ Note, however, that the link between the decision and the action is not guaranteed any more in the first-personal mode than the third-personal mode. The probability that it *does* what it *decides* is secured by the causal/mechanical link between decision and action. Black's practical perspective on the game is a matter of what its decisions control.

⁷ They come together in the discussion of Landauer's principle.

⁸ I have bracketed issues about the Hard problems of consciousness or intentionality by focusing on robots and computers, though I agree with Rovelli that the side-on view that we are taking of the embedded agent representing the world also holds the key to understanding what is going on in the arguments like the knowledge argument.

that are in focus here are ones that we share with any designed artifact or naturally evolved system that uses information to guide behaviour. We can avail ourselves of the side-on view of these systems (the one that spans the level of physics and the level of representation) and talk about how the fact of its own embodiment and the causal link between decision and action show up at the representational level, and also how representational activity ends up being integral to physics. Intelligent processes move bodies, build bridges, dig ditches, and move matter around in ways that don't happen without them. There is interference in both directions. Interference goes from world to representation but also from representation to world.

I also agree with Rovelli that this same strategy should be used to understand what is happening in (for example) the famous Knowledge Argument for the irreducibility of consciousness. The best way to understand conceptual and epistemic gaps is to see how they arise in agents representing the world with several sources of information. I would add to Rovelli's brief remarks the observation that, when we start exchanging information through public channels, objects are stripped of their sensory qualities. Objects get projected out into the physical landscape and sensory qualities are relegated to a private inner world. The reason for this has something to do with the way that sensory qualities vary with who is looking and the setting in which things are viewed, as well as with the impossibility of comparing experience across a sensory divide. It was what led Galileo to distinguish primary from secondary qualities and it seems a crucial part of understanding the irreducibility of phenomenal properties. Note also the connection to Deng's remarks about A and B properties in the temporal case. What Deng says is that A and B concepts get partially and provisionally detached from the more primitive A/B structure for purposes of being able to separate invariant properties from the flux of experience. That is what allows us to think about particular events (characterized as nodes in an abstractly characterized relational network) from different perspectives in that network. This is of a piece with the separation of the conception of self as an object moving through the world from the conception of the world as an object being moved through. This conceptual articulation of experience is what permits the synthesis and storing of information acquired at different times and places and the exchange of information across a perceptual divide. The world is, in this way, discerned in experience. This is not of course, accomplished by any conscious application of reason; it is prior to reasoning, a condition of its possibility, effected by pre-perceptual processing. The only way to bridge the divide is to see them as rooted in the primordial structure which is prior to both. To see physics as an abstraction out of experience and experience as part of the world.⁹

Rovelli writes:

In my earliest attempts to understand Ismael's paper, I overestimated the paradox presented in the opening. Ismael's interference regards the effect of knowledge on the future, whether or not there is anything paradoxical in the situation. Hence it is not the threat of paradoxical situations that opens the future: it is the more general phenomenon of interference. (Rovelli 2024: 263)

Yes, he is exactly right. The point of using the paradox was partly because it is a very nice, concrete embodiment of interference in its purest form. It captures in a clean way what I mean by interference, how it arises and what it entails. It should go without

⁹ This is a point also made in some detail in Frank, Gleiser and Thompson 2024.

saying that there is no inconsistency in reality and no true contradictions or anything like that. What it shows is only that negative interference will keep a system in the world from representing it in complete and accurate detail, where completeness is understood here in a particularly strong way, to include both positive and negative respects (truthful answers to all yes/no questions of fact). The upshot is simply that our representations of the world must recognize and respect interference (and so maintain a kind of perpetual incompleteness) on pain of contradiction.

Putting it in terms of self-reference might make it seem that it is really about representation, but that is not quite correct, for the reasons that I emphasized in the opening remarks. It also reveals something important about physics. To the extent that the prediction of the future from knowledge of the past works by applying laws of temporal evolution to initial states, the computation mirrors the physical evolution of the world, and the self-reference encountered by an embodied system trying to predict the future is a reflection of the fact that its own activity is integral to the physical evolution. The physical process that leads from the past to the future passes through its own deliberation and reveals that deliberation plays an essential and integral role in how the bodies of creatures like us move and hence in the often quite radical changes we make to the physical environments we inhabit. There is no way of bypassing it: no way *to* the future without passing *through* the deliberation. We could have made the same point with positive interference. Just as there are some questions that can't be answered truthfully in the moment, there are also some that have *multiple* truthful answers: 'is the answer that is about to appear on the output screen "yes?"', 'is the output of this deliberation A or B or C?'. In these cases, alethic constraints disappear, and there is no way to calculate through in the descriptive mode. Positive interference produces several consistent solutions and that means that at the point of degeneracy, the prediction becomes a decision.

These are just different ways of making the point that Rovelli and I agree is the important one, *viz.*, the inextricability of physics and representation (with its allied notions of knowledge and truth). No matter where you start—from the point of view of physics, or from the point of view of representation—you are going to come up against the fact that your representational activity cannot be disentangled from the world you are representing.

I don't think this is a problem: rather—like Rovelli—I think it is a solution to a cluster of difficulties that emerge if it is ignored.

Near the end of his commentary Rovelli writes:

many physicists see the apparent clash between determinism and the openness of the future not so much in terms of the knowledge available at some time, but rather in the apparently puzzling fact that in a deterministic theory the data (that can be gathered much later) about a given spacetime region are not compatible with different events in the future causal domain of dependence of that region. (Rovelli 2024: 264)

Yes, I agree. The point about relativity is invoked in the paper to defuse one particular way of interpreting the openness of the future: one that draws a distinction between nomological openness and epistemic openness, holds that the future is ontically fixed by the past, but is epistemically open because some of the information that is relevant to its nomological fixation lies outside of our view. That idea depends quite specifically on imposing an asymmetry on events in the absolute elsewhere that relativity doesn't recognize. In a Newtonian setting, there isn't a way to hold that the

past is fixed, and the future is open except in a merely epistemic sense because—if the laws are deterministic—the former entails the latter as a matter of physical necessity. Things are different in a relativistic setting. Relativity brings temporal structure into alignment with causal structure, imposes a finite speed on the propagation of influence and information, and gives us a local account of physical processes. In so doing, it tells us that the operative physical constraints on what happens at any point are contained in its past light cone and it also changes the logic of our temporal categories. What Rovelli (in other places) calls ‘the grammar’ of past and future is different in a relativistic context. Temporal relations are no longer relativized to a moment in time, but instead to a point in space-time: ‘now’ becomes ‘here-now’. Instead of past and future, there are now: causal past, causal future, and absolute elsewhere. The logic of the relations between these three categories is given by the Minkowski geometry. If we look at what determinism entails in this setting, the old argument that used determinism to leverage the fixity of the past into the fixity of the future no longer works. In the Minkowski geometry ‘past is fixed, future is open’ is perfectly consistent and we can’t assimilate events in the absolute elsewhere to either category (see Ismael 2023a).

Rovelli’s remarks call attention to the fact that even if we can recover a kind of prospective openness in a relativistic setting, it remains true that from a retrospective perspective—if, for example, at the moment of my death, I were to take a cross section of my past light cone from the distant past (long before I was a gleam in my mother’s eye)—the data along that cross section would fix everything that would happen to me and everything I would ever do. That thought seems upsetting; it seems to undermine the openness of the future from a prospective perspective. You are supposed to look back and think ‘oh, I see now retrospectively that my actions were already determined even while I tossed and turned, even at the moment that I stood on the threshold of the decision. I see now that all of these events in the distant past were already on their way to cross into my light cone.’ You need some such thought for the retrospective perspective to undermine the prospective perspective. But that is a relativistically illegitimate thought. The only way to get the retrospective perspective to undermine the prospective openness, is to implicitly impose the old logic of past and future on the relationship between perspectives. There is no good relativistically legitimate meaning to ‘already’ in the illegitimate thought. In a relativistic world the retrospective perspective doesn’t undermine the prospective one. The logic of past and future is just different in this setting. Looking forward, the future is completely open; looking back, it seems like it couldn’t have gone any other way. The puzzling fact remains just that: a puzzling fact. such is the perverse logic of relativity.

Shrapnel, Evans and Milburn

While most of the commentaries focus on how interference plays out at the representational level, Shrapnel, Evans and Milburn focus on physics. They introduce a physical model of a minimal system that models the world and uses what it learns to guide its behaviour.

They write:

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 world. (Shrapnel, Evans and Milburn, 2024: 279)

The beauty of this device is that it is a physical embodiment of the relations between the level of representation and the level of physics. Each of these levels has a temporal arrow and the possibility of getting a general characterization of relations between them is due to the centrality of thermodynamics to both. The physics of this device gives us a physical understanding of the alignment of the arrows. The insight conveyed by their model is that the alignment of arrows is not a conceptual matter, but a matter of physics.

Their toy model captures, moreover, the mutual dependence of the agent's internal representation of the world and the world in which the agent is contained. The agent and its representational activity play an integral role in the physics of the world it is representing and the agent's model develops in response to, and is meant to mirror, the causal structure of the world. Interference is the manifestation of that mutual dependence.

Evans, Millburn and Shrapnel appreciate that it is because the agent's representational activity plays an integral role in the physics of the world that it is representing that the world is incomplete without it. The openness, for that reason, is not merely epistemic. The agent's representational activity plays an ongoing role in the world as it develops. The world is, in fact, coming into being as it is experienced by the agent.

The two additional papers Shrapnel, Evans and Milburn cite (Evans Milburn and Shrapnel 2021, 2023), in which their model is introduced are full of rich and illuminating discussion. I recommend them highly as a beautiful example of how philosophical questions can be elucidated by physics. I rely on them in my own thinking about these questions.

Adlam

Adlam notes the individualistic focus of both mine and Price's accounts of the practical asymmetry and suggests that the thermodynamic gradient may not guarantee either that an individual agent has the same temporal orientation over the course of her life or the alignment of the internal arrows of agents spread through spacetime. She rightly notes that self-reference is a limiting case of closed causal loops. She introduces a consistency in chaining requirement which says that the output of a process oriented in a certain temporal direction can only be used as the input to other processes oriented in the same temporal direction. This requirement keeps us from chaining together processes into loops on the grounds that such loops would create logical inconsistencies.

Here is what she has in mind:

suppose process P takes as input a bit and its effect is simply to flip the value of that bit, and suppose we perform the chaining such that the input of process P is equal to its output. Then we end up with a logical contradiction: if the original input to process P is bit value 0, then the output must be 1, but by construction the output is equal to the input, so there is no logically consistent way of resolving this situation. Thus, if the universe cannot contain logical contradictions, it must be impossible to chain processes in this way. (Adlam 2024: 286)

I don't think I agree that there are no logically consistent solutions except in the special case of a closed, deterministic process. As soon as we embed a system in the wider

world and allow exogenous influences that can break the chain leading from output back to input (any physically possible event no matter how improbable—for example, a meteor from outer space or a banana peel in the wrong place) we have consistent solutions. That is all that is needed to avoid logical contradiction. Processes in the actual world are never completely isolated. We string a bunch of events together in a way that is intended to produce an effect, but no matter how carefully we try to shield them there are always potential exogenous influences and that means there will always be a consistent solution in which some exogenous influence foils the process and keeps it from completing as intended. This is the Wheeler/Feynman (1949) resolution of the paradoxes associated with closed causal loops. The upshot is that in a world where the consistency in chaining constraints were violated, I would think that we would find physical anomalies (physical correlations and oddities that we don't see in their absence), but not logical contradictions (Ismael 2003).

I've explored these issues in another paper, where instead of a closed causal loop I have a Laplacian demon predicting the outcome of a counter predictive device (Ismael 2019a). The structure is the same. The Laplacian demon's prediction acts there as a proxy for the output of the device since in a deterministic world a Laplacian demon predicts that p iff p . There I use the argument to explore whether classical mechanics is caught in something like the inconsistencies that she thinks arise when consistency in chaining is violated. I argue that it is not in that case because the possibility of embedded demons of this kind are not in fact provided for in classical mechanics so classical mechanics in fact satisfies the consistency in chaining requirements. But even if classical mechanics provided a mechanism for perfect predictors, I think these constraints are weaker than Adlam does (though I'd be very keen to have her response). They aren't logically binding so long as there are consistent solutions that incorporate them and so long as we are talking about chaining together the inputs and outputs of open systems, there will be.

I don't think that this affects her main point. I love what she does in the commentary. I agree that interference plus promiscuous chaining (any two physical processes can be chained together to produce a third) leads to a global order with a consistent direction. And that it can solve the sort of global coordination problem that appears in the classical case and is especially urgent for the quantum reference frames program.

Price

Price argues (i) that we don't really have 'the agential viewpoint' until we have an agent that is deliberating about what to do, (ii) that once we have the agential viewpoint, there's an internal arrow that runs from deliberation to action and 'what we really mean by "the future"' is the direction in which we deliberate, and (iii) the reason that we think that we can affect the future but not the past is a matter of definition.

On (i): when Price asks: 'Is a self-referential observer necessarily an agent?', the answer is that it depends on what we mean by 'agent'. We can use 'agential viewpoint' to mean the full deliberating, decision-making viewpoint, but there's a weaker sense in which the 'agential viewpoint' is forced on anything that falls in the scope of its own representation and the richer sort of agential viewpoint is built up around it. In the thinnest sense, anything that exists is an agent because some of what happens are things that it does. This goes for a rock, a dust particle, a butterfly, or a tree. It is the strategic use of agency in that thin sense, together with the connectedness of

one's actions in the wider environment that biology exploits and intelligence perfects. Price is using 'agent' in the thick sense to refer to a deliberating agent and arguing that a self-referential observer is not an agent in that sense. I agree. To some extent this is a matter of terminology. I prefer to start with the thin, unavoidable type of agency, then build it up in stages by adding connectedness and the strategic exploitation of agency because it lets us see how agency in the thick sense arises by steps that are individually unmysterious. That takes away some of the mystery that surrounds the emergence of agents. In the process as I describe it, crossing the threshold from observation to agency is made at the first step in an unobjectionable way and exploited in increasingly strategic ways as we move through the stages.¹⁰

So, in my terminology, agency (in the thin sense) plus the connectedness of one's actions in the wider landscape leads to interference. Deliberation is the strategic use of interference to affect the probabilities of events in the wider landscape. This way of putting things also brings into relief the conditions that support agency in the thick sense. Agency in the thick sense requires both external scaffolding and internal (cognitive) complexity. It requires the flow of information not just through perceptual pathways, but also out into the landscape via its effect on movement. It's the connection between thought and action, together with the possibility of predicting the effects of potential actions, that opens up the space for deliberation. So, it was self-reference against the background of the physics we have, including the thermodynamic gradient and all of the scaffolding supports the everyday world, that I suggest that self-reference leads to decision.

So, agency in the thinnest sense is there by default; but I agree with Price fully that deliberative decision-making involves a good deal more than simply self-reference. There is the external connectedness of one's actions in the wider landscape that was missing for his example of the self-referential angels, and there is the internal functional organization that is organized around trying to harness interference for one's own ends that is missing for rocks and planets and tornadoes. Even with all of that in place, one can simply decline to take up the reins, letting the world pass like a disinterested observer: even with respect to thoughts and emotions, meditation experts sometimes advise that we should observe them, acknowledge them, and let them pass without judgment or an attempt to control.

On the question of 'what we really mean' by 'the future', I am agnostic. I'm happy to acknowledge that there are distinguishable arrows and address questions about how they align. Price's way of arguing for the claim that what we really mean by 'the future' is the direction in which we deliberate is to entertain a thought experiment in which we take a fully formed agent with its own internal deliberative arrow and imagine that we place it in the world in such a way that it was reversed with respect to the thermodynamic gradient. In that case, he thinks that the agent would see themselves as living in a world with a high entropy past and low entropy future. He wants to

¹⁰ Something similar can be said about information. The thinnest most basic notion of information is mutual information. This is directly definable in physical terms. If you start from there and see how to build up to the richer intentional notions in steps, by adding scaffolding and structure—some of it contributed by the agents that use those notions and whose content is partly given by the inferential network in which they are connected—the path looks intelligible. If you start with the rich concept and try to reduce it directly to physics, it won't work. The concepts of physics are not rich enough for a direct reduction, but they are rich enough for building the context in which those notions arise. I believe a similar story holds for causation. You don't try to reduce concepts to physics; you try to build minds out of physics that use concepts. Concepts are then understood functionally by their role in minds.

say that what the future means to us is the direction of deliberation, so if we were taken out of this world and placed in an environment in which our internal arrow did not align with the external one, for us the future would still be the direction in which we deliberate. Maybe. But I'm not sure what hinges on it.

When I first read Price's commentary, I didn't see the point of trying to imagine a transcendent agent plugging into the world from the outside. To me it seemed rather giving too much pride of place to the internal arrow, which should be seen rather as the last piece in a puzzle, a piece whose shape is determined by natural opportunities for agency. It goes against the whole spirit of the approach here which is to embed the agent in the world and treat it fully as part of physics. Where Price wants to emphasize the *separability in the imagination* of these arrows, I want to emphasize their *physical connectedness*. We treat agents not only as themselves physical systems (hence subject to the laws) but evolved to take advantage of opportunities that physics provides. As soon as we embed the agent in the world and treat it as part of physics, we get a natural agent with an internal arrow aligned with the thermodynamic gradient for two reasons: (i) the one emphasized by Shrapnel, Evans and Milburn: *viz.*, any physically implemented IGUS (information gathering and utilizing system) is going to have its internal arrow aligned with the thermodynamic gradient because it will be a dissipative system, and (ii) the one emphasized by the evolutionary story: because those are the natural opportunities that the universe provides for agency in Price's rich sense. Deliberative agency evolved to exploit the macroscopic information made available (in part) by the thermodynamic gradient in the form of records.¹¹ To my mind, the illuminating perspective is not unconstrained imaginings that treat the agent as a transcendental system with a *sui generis* mental life and plug it into the world by making choices about what it has perceptual access to and what it controls like a transcendental Maxwellian demon intervening from the outside. This can make our internal perspective seem self-supporting and unrooted. It is rather the side-on perspective that reveals the physical connections between the arrows. The connection between the arrows is physical not conceptual. And the explanatory direction goes from the past hypothesis and thermodynamic gradient to our internal perspective.

On reflection, though, I do see the point of Price's emphasis.¹² First, he wants to discourage the tendency to reify the internal arrow in the external world. And he wants to pave the way for retrocausal interpretations of quantum mechanics.¹³ So he wants to say 'look, the difference between past and future as you experience it is a difference in the direction in which the probabilistic effects of interventions propagate. Classically that is in the future direction, but at the quantum level, influence runs in both directions. With these, I agree.

Hoerl

I found Hoerl's (2024) discussion intriguing but also puzzling because his classification is foreign to me and I would have disagreed with the way he places my own position. He distinguishes empiricist from rationalist conceptions of the flow of time: 'empiricist

¹¹ See Ismael 2023b for closer discussion of these sorts of thought experiments and some of the associated technical questions that we don't have room for here.

¹² He set me straight in personal correspondence.

¹³ I don't think you need to reify it; you just take seriously enough that it is embedded in, and hence part of the external world.

approaches to the idea of the flow of time ... assume that conscious experience presents us with (apparent) empirical evidence of the flow of time' (253). Whereas rationalist approaches hold that '[i]t seems impossible to conceive of a conscious point of view on the world in the first place that does not involve an awareness of the passing of time' (253). He says that he finds elements of rationalism in my view, and (at one point) that I argue that 'thinking about the flow of time in terms of the idea of Becoming is really part of what it is to think of oneself as an agent in the first place' (Hoerl 2024: 257).

I wonder why he says that. If you had asked me 'is it possible to conceive of a conscious point of view that does not involve an awareness of the passing of time?' I would have said 'yes'. I assume that we are conscious, but I have no views about what is constitutive of consciousness or whether it requires a conception of oneself as temporally extended. I don't see a clear argument that there couldn't be a phenomenally conscious being that didn't have a concept of the flow of time. I suspect that he has some other notion of consciousness in mind, perhaps one more closely linked to the idea of a first-person point of view on the world where that includes a sense of diachronic identity and something like Price's full agential point of view. If that is what he means, then I do think that if you an information gathering system that explicitly represents time and stores large bodies of temporal information to use strategically in the service of action, you cannot think of the future as there to be known.

Near the end of his comments Hoerl writes:

if ... thinking about the flow of time in terms of the idea of Becoming is really part of what it is to think of oneself as an agent in the first place, then I think this runs counter to the idea that our common sense understanding of time can simply be unmasked as a matter of perspective. For if it was just a matter of perspective, there would have to be a way of transforming that perspectival view of time into what Bernard Williams (1978) has called an absolute conception of reality, resolving differences between particular points of view. (Hoerl 2024: 257)

I don't know why he says this would be a matter of 'unmasking', and this may be a place where Deng's commentary is helpful. Common sense recognizes that notions of past and future are perspectival in the sense that they are relativized to a moment.¹⁴ We all know that today's tomorrow is tomorrow's today, that the present will tomorrow be the past, and so on. Fixity and openness, insofar as they attach to events in virtue of their status as future, are then also of course perspectival. Consider some future event: for example, what I eat for breakfast Jan. 1, 2026. That is open from a prospective point of view and fixed from a retrospective point of view, so whether an event is fixed or open is relative to a temporal perspective.¹⁵ This is just part of the ordinary logic of past and future.

On the question of whether there is a way of transforming the view of time through the eyes of the embedded human being into an absolute conception of reality, resolving differences between particular points of view: in one sense yes. I think we can certainly resolve the differences between temporal perspectives as easily and smoothly as spatial ones. You and I know how to understand 'oh, I was so looking forward to the concert

¹⁴ Perspectival, here, means implicitly relativized.

¹⁵ Relativity does teach us that there is an additional element of perspectivalism, one that common sense doesn't recognize: *viz.*, that they are relative not just to a moment in time but to a point along a world line. That, however, plays no role in what I say here, which is restricted to the experience of an individual agent. Relativity adds an asynchrony between the world-lines of spatially separated agents.

in the five months before it happened, but the last month got so stressful that, when it was actually happening, I was tense from beginning to end. Now, I'm just relieved it went okay and I can enjoy the memories.' The ability to represent events in our own lives in anticipation and then in retrospect and to relate them extensionally through our different temporal relations to them is a normal part of the human understanding of time.

And I disagree with him that the idea of Becoming that I have argued for is opposed to the availability of an absolute conception because, as he says: 'it implies that there is something about things in time that itself changes over time', (Hoerl 2024: 257).

The only thing that changes in time is the agent's perspective and it changes only in the sense that it is different at different times. There is nothing that changes in the absolute sense, but if we compare the agent's perspective at different times, it is different. The agent represents the same events over and over through the course of her history, first in prospect, then as they happen, and then in retrospect. The change she 'sees' is the change in her perspective.

There is, however, another idea that is sometimes included in the idea of an absolute conception of the world: the idea that an absolute conception of the world would be a complete and accurate catalogue of the totality of facts. When we try to stabilize an absolute conception of the world in that sense, we encounter interference. We can't pretend that the universe in that sense is *there* to be referred to de re from our position in it, well-defined in every detail, independently of what we say or believe about it, adjudicating fact from fiction. The part of Hoerl's discussion where he seems to be talking about this, I really like, and would like to understand how the other parts connect to this idea. He writes:

Our status as 'embedded' agents which Ismael is so keen to emphasize makes it impossible for us to adopt the kind of detached point of view this would require, from which we could compare our representations of the world with reality itself.⁷ As Ismael herself says, "it is customary in cosmology to maintain the imaginative fiction that we are sitting outside the universe looking down" (2024: 216)). It is exactly that, a fiction. (Hoerl 2024: 257)

Yes.

Braddon-Mitchell and Miller

Braddon-Mitchell and Miller have two complaints:

1. The point made in the paper about special relativity isn't sufficient for freedom.¹⁶
2. When it comes to temporal experience, interference is something but not everything.

On 1, I agree. I wasn't arguing otherwise. The word 'freedom' pointedly does not occur in the body of the paper (it is mentioned, but not used in the references as part of a title of a book).¹⁷ The point about relativity was invoked here to make a technical point against an epistemic reading of the openness of the future on the grounds that the laws entail that the past nomologically fixes the future, that is, that the past

¹⁶ This is a point made more carefully in Ismael 2023a.

¹⁷ Ismael 2019b. Indeed, to see just how much I agree, see the book.

fixes the future as a matter of nomological necessity, no matter whether we are in practice ignorant of it or not.

On 2, yes, of course; I also agree about that.

I also like their suggestion that there might be a ‘correlation between the ability to regard one’s own deliberation and psychological processes dispassionately, or at least in a way that doesn’t feed into the process being monitored and predicted, with reported differences in temporal phenomenology’ (2024: 250) and the idea of investigating that empirically.

Places where there is some disagreement is that they seem a little dismissive of interference as having anything illuminating to say about the openness of the future when they write: ‘we are just left with the observation that a deliberative process cannot itself predict what it’ll decide other than by deciding. But other processes can’ (2024: 249). I agree that processes that bypass deliberation can take an epistemic attitude towards them and might be able to predict decisions if the predictions are kept from decision processes, but that doesn’t undermine openness. The crucial point is that I can act on any information that I have, including any attempt to predict what I will do, if I so choose. *That* is both what keeps me from stabilizing the future and makes my decision the last word. The fact that other people or other processes have to keep their predictions secret to avoid getting undermined is exactly the sense in which the say is ultimately mine; the direction of determination goes from what I do to whether their prediction is right. The prediction is beholden to the action rather than the other way around. This point is neither more nor less than it is. It is entirely compatible with determinism, entirely compatible with predictability by non-interfering bystanders, and entirely compatible with predictability by myself in advance, if I choose to do what I predict.

The mere fact of predictability has never, in any case, been undermining to the openness of the future or one’s power to do any of a number of things. My husband can predict what I will order in a restaurant. And I can predict myself in all kinds of ways. That doesn’t mean that the choice is fixed in advance, or that I’m not free to make whatever choice I want. What does it mean to be free to make any choice I want? I’m not sure most people outside the philosophy classroom have an articulate idea, but one thing it involves is that nothing forces me, and you try to predict what I will do, I can do otherwise. If you doubt it, tell me your prediction and let’s see who wins that game. Take as much time and avail yourself of any information you like, and let’s see what happens. In any direct confrontation between your prediction and my action (where by ‘direct confrontation’ I mean that I am made aware of your prediction), I can thumb my nose and there is no physics argument to the contrary (see Ismael 2019a). The slip from ‘predictable from knowable facts that were in place in advance with certainty’ to ‘couldn’t do otherwise’ was supposed to link predictability to freedom. And the argument for ‘predictability in advance with certainty’ was supposed to come from determinism. The line of argument here breaks that link. It slips right into that crack. Of course, lots of other things remain true that you might think point in the other direction. I’m not going to act in a way that is contrary to the laws or that can’t be seen when all is said and done as entailed by data along any Cauchy surface. But don’t tell me that there is any Cauchy surface that falls in my past light cone, and whose status as a Cauchy surface is determined by data in my past light cone. At the time I act, my action isn’t already fixed in the meaning of ‘already’ that is applicable here. It is precisely the compatibility of these things that

forces us to do a little soul-searching about this whole cluster of concepts. There's a razor edge of consistency that relativity shows us how to walk between global determinism and the local openness of the future that hasn't been emphasized enough in the literature. I see no way to get the global determinism to undermine the local openness without imposing an asymmetry that relativity doesn't recognize.

Make of that what you will, but you aren't going to be able to use determinism to argue that while the future is open epistemically it is nomologically fixed. My own view about this is that instead of going back and forth in the fight between freedom and determinism, we should recognize that neither freedom nor determinism are what we thought they were. There is what there is, and there isn't what there isn't. To some extent the old opposition has been overcome. This is another place where I think Deng's way of putting things is really helpful. To understand what freedom really is for us, we need to go back to a way of seeing things before the philosophical positions ossified intuitions that aren't part of common sense. The weakening of determinism in a relativistic context makes room for some kind of freedom, some kind of control over our destinies, but to understand what it amounts to, one has to understand life and cognition and nothing less than the human being in all of its physical complexity.

Zhu

Zhu argues that the account of openness in the paper involves an unjustified 'slide' from 'instability of representations to representations of instability' (2024: 266). He argues that the assumption should be rejected and moreover that accepting it threatens to undermine our conception of the past as fixed.

I disagree. There wasn't an unjustified 'slide' from representations to the world, but rather a denial that the separation could be ultimately maintained. The central claim of the paper is that we can't quarantine instability to the level of representation because the distinction between representation and represented breaks down. Your representational activity is part of the world, integral to what happens, and entangled with events in the wider landscape by way of its effects on the movements of your body. Interference is the manifestation of that entanglement.

Zhu asks: 'Why cannot I simply think that I have contrary views at different times about the same, fixed piece of reality?' (2024: 267) The answer is of course, you can have contrary views at different times about any 'piece of reality'. But you can't treat the 'piece of reality' as 'fixed' independently of your representation of it when the 'piece of reality' is your representation, or anything that is physically dependent on it so that they vary together. That is the difference between ignorance and interference.

Zhu thinks that the move from instability of representations to instability of facts leads to anti-realism about the past. Again, I disagree. The difference between the two—which the thermodynamic account of the epistemic asymmetry was invoked in the paper to underwrite—is that there is something that breaks the probabilistic dependence of the past on present action.

Consider two cases:

- (i) You are a detective trying to solve a crime. When you first approach the scene, you think it is a robbery gone wrong. On day 2 you begin to suspect the husband. On day 3, the butler becomes your prime suspect. You change your

mind several more times over the course of the investigation. You want to crack the case. You know there is an answer, and you are bent to find out, but you waver in what you think it is.

- (ii) You have a potentially life-changing medical decision to make. You have to choose between a lifetime of medicine to manage a chronic condition or a risky procedure that might cure it. You toss and turn all night, thinking ‘I’ll do this, no I’ll do that’.

In both cases, one’s representation of a fact is changing over time.¹⁸ The first is a case of ignorance; the second is a case of interference. What is the difference? In the first, the fact that you are representing (who committed the crime) can be stabilized independently of your representation. In the second the fact that you are representing (whether you will get the operation or not) is tied to your wavering, and hinges on what you decide.

Physically what underwrites that difference? Why does interference infect future events but not past ones? Here is a sketch. To investigate how our own representational activity is connected in the world it represents, we assume that our beliefs affect our actions, so the question is whether anything I believe or do in the here and now is going to affect what happened in the past. To answer that question, we take the present surveyable macrostate of the world, use the microcanonical probability distribution to generate a probability distribution over possible microstates compatible with that macrostate, and evolve it backward, conditionalizing on the past hypothesis and the decision to perform some voluntary act A. Then we compare that with what we get by performing the same procedure, conditionalized on decisions to perform alternative voluntary actions B, C, and D. It doesn’t matter what information I actually possess, so long as there is some record (that is, some potential macroscopic trace, no matter how insignificant) of a past event, this will shield it from probabilistic dependence on A. Conditionalizing on the past hypothesis screens the past off from probabilistic effects of what I do or believe in the here and now and provides us a quite general way of protecting past events from interference.

This sketch is intended to convey the contours of an account whose details are still being sorted out. It relies on a well-established, but ongoing discussion in the physical and philosophical literature about how to understand questions like:¹⁹ What makes a footprint in the sand a record of a walk across the beach? What makes a scar on a tree a record of past injury, or a burnt forest a record of a fire? What is the physics that

¹⁸ ‘Stability’ in the relevant sense is not stability simpliciter (unvarying over time). It is both relational and time indexed. The relata are events (the temporally localized ‘pieces of reality’ which are the basic elements in a relativistic world) and representational acts. Whether an event can be stabilized independently of a representational act depends on how the two are related. Perhaps he has in mind an alternative, atemporal notion of fixity according to which everything that happens is fixed. I’m afraid that I don’t see what it means to say that everything that happens is fixed, unless it means that everything that happens happens. And even if such a notion is granted, it doesn’t stand against the fact that for me, in the here and now, there are all kinds of facts about the world that are not fixed already and that can’t be stabilized (in my sense) independently of my representational activity, and therefore cannot be treated as there to be known. And the world as a whole (the totality of facts, past, present and future) is never as such, for me, a fixed object.

¹⁹ See, for example, Reichenbach 1999; Albert 2015, Ch 5, 2000, ch. 6; Rovelli 2018, 2019, 2022; and Fernandes 2022. There are some differences, noted in the paper, between my sketch and that of Albert and Fernandes: since my sketch deals with probabilistic dependence and not counterfactuals it is less vulnerable to the kinds of counterexamples that set up a fragile counterfactual dependence between present action and past fact of the kind that their account is susceptible to.

underwrites the informational properties of this diverse class of physical processes? How are they connected to the Past Hypothesis? Do they provide a complete barrier to causal influence propagating into the past, or only an approximate and emergent barrier that keeps in influence from propagating back in time for all practical purposes?

Zhu worries that this account leads to anti-realism about the past. He writes:

There is no guarantee that the traces and records available at one time will be the same as the traces and records available at another. The traces and records that are currently available might become unavailable at a later time (for example, they might be changed or destroyed, etc.) and those that are currently unavailable might become available later (for example, new traces and records might be discovered; new technologies of accessing traces and records might be invented, etc.). (2024: 268)

In saying this, he is relying on a notion of ‘availability’ that is more anthropomorphic than the one that is relevant to the sketch. So long as there is some potential information encoded in the present state of the world that could be—by a physically possible process—accessed by a measurement and registered in the position of a pointer observable, there is a trace or record available in the sense relevant to the sketch. The sketch did not say anything about what kinds of information we might have preserved in our museums, hidden in our drawers, or registered in our history books. It didn’t say anything about what instruments or technologies we might have in our toolkits or whether our theories were up to the job. Even for questions that we don’t have the answer to and don’t know how to settle with present methods, it makes sense to treat the unsettledness as ignorance and wait until methods become available, precisely because the relevant kind of availability is ‘accessibility by a physically possible process’. That is what makes it appropriate to take an epistemic attitude to the past, that is, to look for evidence to settle factual questions about the past and to treat our lack of knowledge as ignorance.

Zhu does raise a good point, however, which is that the sketch certainly needs to be filled out to deliver a full understanding of human practices of knowledge gathering about the past and to make the role of the thermodynamic gradient clear once we move past fossils and footprints. We need to make more explicit, for example, that human practices of learning about the past are not reliant on whatever records have been passively registered by natural methods in the world’s macrostate. They involve accessing information buried below ground, using techniques like radiometric dating methods to determine the ages of rocks or fossils, and telescopes to detect microwave background radiation. We need to make more explicit the fact that measurements (which transform information encoded in the microstate of a system into macroscopic form) are ways of producing records and themselves reliant on the thermodynamic gradient.²⁰ The important part of all of this for our purposes is that whenever we can *find* or *produce* a record of a past event, that will screen off any probabilistic dependence on present action and that is what makes it possible and appropriate to take an epistemic attitude towards the past, i.e., to treat whatever lack of knowledge we have as ignorance in a classical setting.

It is different for the future. Future events—not all future events, but many of those that are in the foreground of our experience, and in ways that we can’t precisely

²⁰ This connection was made explicit in Albert (2000), *ibid.* A measurement is a process that correlates a (typically unobservable) feature of the world with the macroscopic feature of a measuring device) in such a way that the state of the pointer after the interaction is a record of the state of the system before. The thermodynamic gradient is relevant because the inference from the final state of the device to the initial state of the system depends on the Past Hypothesis as surely as the inference from a footprint in the sand to a passerby.

contain or control— are afflicted by interference. Interference is, in my view, what was needed to make sense of the stance or attitude or presumption—the deeply felt, pre-cognitive conviction—that the future is not yet locked in; i.e., that while you are wavering over a difficult decision, the future itself (not just your representation of it) hangs in the balance.²¹ That is what the paper was about. I was taking the epistemic asymmetry for granted and adding something to it (interference) that would give explicit literal content to that specifically temporal sense of unfixity, and to capture the elusive difference between time and space that I conveyed with the landscape analogy in the first paper when I said we think of the landscape as a fixed object coming into view but the future as coming into being as we experience it.

The kind of instability I'm talking about isn't mysterious. It doesn't threaten objectivity or realism; it doesn't even have to do with truth. The objective facts are the objective facts. What is true is what is true. It is just that our own knowledge (and knowledge gathering activity) is connected in the world in a way that keeps us from treating the world as a pure object of epistemic attention. This is what it is to *exist*, for your life to *part of the world*, for the world to be coming into being as you experience it, unfolding in real time, *waiting to be decided*.

References

- Adlam, Emily (2024) 'Self-Reference, Time-Asymmetry and the Page-Wootters Formalism', *Australasian Philosophical Review* 8: 285–290.
- Albert, David Z (2000) *Time, and Chance*. Harvard University Press.
- Albert, David Z (2015) 'The Difference between the Past and the Future', in *After Physics*: 31–70. Harvard University Press.
- Braddon-Mitchell, David, and Kristie Miller (2024) 'On Explaining Temporally Asymmetric Experiences', *Australasian Philosophical Review* 8: 243–251.
- Deng, Natalja (2024) 'Self-reference, tenseless passage, and squaring the (philosophy of) time circle', *Australasian Philosophical Review* 8: 272–277.
- 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/.
- Fernandes, Alison (2022) 'How to Explain the Direction of Time', *Synthese* 200: 389. doi: [10.1007/s11229-022-03818-4](https://doi.org/10.1007/s11229-022-03818-4)
- Frank, Adam, Marcelo Gleiser and Evan Thompson (2024) *The Blind Spot*. MIT Press.
- Hoerl, Christoph (2024) 'The Flow of time: Rationalism vs. empiricism', *Australasian Philosophical Review* 8: 252–259.
- Ismael, Jennan (2003) 'Closed Causal Loops and the Bilking Argument', *Synthese* 136: 305–320.
- Ismael, Jennan (2019a) 'Determinism, Counter predictive Devices, and the Impossibility of Laplacean Intelligences', *Monist* 102: 478–498.
- Ismael Jennan (2019b) *How Physics Makes Us Free*, Oxford University Press.
- Ismael, Jennan (2023a) 'Rethinking Time and Determinism: what happens to determinism when you take relativity seriously', in Remy Lestienne and Paul A Harris, eds., *Time and Science, Volume 1: The Metaphysics of Time and Its Evolution*: 147–172. World Scientific Publishing.
- Ismael, Jennan (2023b) 'Reflections on the asymmetry of causation', *Interface Focus*. 13: 2022008120220081. doi: [10.1098/rsfs.2022.0081](https://doi.org/10.1098/rsfs.2022.0081).
- Price, Huw (1988) *Facts and the Function of Truth*. Blackwell.

²¹ 'Locked in' is not (for me) a transcendental notion, but an embedded one. It is a reflection of the fact that your thoughts and decisions in the here and now are putting in place things that are integral to the physical landscape, critical to some of what happens later. I don't think that there is a coherent transcendental notion of fixity that does reify time as an external dimension.

- Price, Huw (1992) 'The Direction of Causation: Ramsey's Ultimate Contingency', *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association, Volume Two: Symposia and Invited Papers*: 253–267.
- Price, Huw (1996) *Time's Arrow and Archimedes' Point: New Directions for the Physics of Time*. Oxford University Press.
- Price, Huw (2011) *Naturalism Without Mirrors*. Oxford University Press.
- Price, Huw (2013) *Expressivism, Pragmatism and Representationalism*. Cambridge University Press.
- Price, Huw (2024) 'The Practical Arrow', *Australasian Philosophical Review* 8: 234–242.
- Reichenbach, Hans (1999) *The direction of time*. Dover Publications.
- Rovelli, Carlo (2018) *The Order of Time*. Riverhead.
- Rovelli, Carlo (2019) 'Where was past low entropy?', *Entropy* 21: 466. doi: [10.3390/e21050466](https://doi.org/10.3390/e21050466)
- Rovelli, Carlo (2022) 'Memory and Entropy' *Entropy* 24: 1022. doi: [10.3390/e24081022](https://doi.org/10.3390/e24081022).
- Rovelli, Carlo (2024) 'Comment on Jenann Ismael's "The Open Universe: Totality, Self-reference and Time"', *Australasian Philosophical Review* 8: 260–265.
- Shrapnel, Sally, Peter W Evans, and Gerard J Milburn (2024) 'Stepping down from mere appearance: Modelling the 'actuality' of time', *Australasian Philosophical Review* 8: 278–284.
- Weinberger, Naftali, Porter Williams and James Woodward (2023) *The Worldly Infrastructure of Causation*. <https://philsci-archive.pitt.edu/22125/>
- Wheeler, John Archibald and Richard Phillips Feynman (1949) 'Classical Electrodynamics in Terms of Direct Interparticle Action', *Reviews of Modern Physics* 21: 425–33. doi: [RevModPhys.21.425](https://doi.org/10.1103/RevModPhys.21.425).
- Zhu, Zijian (2024) 'The Openness of the Future and the Instability of Representation', *Australasian Philosophical Review* 8: 266–271.