AUSTRALASIAN PHILOSOPHICAL REVIEW
2024, VOL. 8, NO. 3, 243–251
https://doi.org/10.1080/24740500.2023.2170258

On Explaining Temporally Asymmetric Experiences

David Braddon-Mitchell and Kristie Miller

Department of Philosophy, The University of Sydney and Centre for Time

ABSTRACT

Ismael aims for an understanding of the nature of an embedded perspective of agents in a world. If successful, this would explain a cluster of ways in which from an embedded perspective, we experience the world in an array of temporally asymmetric ways. Moreover, these are ways that have led many philosophers to rather metaphysically inflationary views about the nature of time, according to which time itself really is dynamical, and is characterized by the movement of an objectively (i.e., non-perspectival) present. Ismael aims to explain these features of our experience without positing any such metaphysical picture. She argues that there are constraints which prevent us from taking this Olympian conception of the world and our place in it, and that these, jointly, explain why we experience the world in these temporally asymmetric ways. We take up two related questions. First, what it would mean to say that these constraints are not merely epistemic, and second, how far these constraints get us in explaining why we experience the world in these ways.

ARTICLE HISTORY Received 23 November 2022; Revised 3 January 2023; Accepted 15 January 2023

1. Introduction

In this rich and enthralling paper, Ismael brings together a number of ideas around the way in which our sense of ourselves as agents in time is shaped by the physics of our universe. It aims for an understanding of the nature of an embedded perspective of agents in a world, as opposed to the Olympian extra-universal perspective that we might try to take on the world when doing physics or philosophy, whilst at the same time contemplating the world from a first person perspective. If successful, such a project would explain a cluster of ways in which from an embedded perspective, we experience the world in an array of temporally asymmetric ways. It would, for instance, promise to explain certain aspects of our temporal phenomenology, or at least apparent temporal phenomenology[1]: the fact that it seems as though time has a direction; the fact that it seems as though time flows or passes or is generally

CONTACT Kristie Miller Kristie.miller@sydney.edu.au

[1] We say apparent here, since at least in some cases it is disputed whether there is indeed any such phenomenology.

This is especially true when we consider the phenomenology of flow or temporal passage. While many B-theorists have supposed there to be such a phenomenology (see Dainton 2011, 2012) Le Poidevin (2007) and Paul (2010), plenty have denied this, including Braddon-Mitchell (2013) Hoerl (2014), Prosser (2016) Miller, Holcombe and Latham (2020), Sattig (2019) Miller (2019), Deng (2013) and Bardon (2013). Of course, all these latter authors agree that in some deflationary sense, one that is consistent with our world lacking robust passage, we do indeed have a phenomenology as of flow.

dynamical, the fact that the future seems to be open in a way that the past does not, the fact that we can causally intervene on the future in a way that we cannot on the past, the fact that we have records, especially memories, of the past, but not the future, and the fact that we are agentively directed towards the future in a way that we are not towards the past. The presence of this package of phenomenologies has led many philosophers to rather metaphysically inflationary views about the nature of time, according to which time itself really is dynamical, and is characterized by the movement of an objectively (i.e., non-perspectival) present. Neither we nor Ismael find such a picture amenable. So, Ismael aims to explain these features of our experience without positing any such metaphysical picture. She argues that there are constraints which prevent us from taking this Olympian conception of the world and our place in it, and that these, jointly, explain why we experience the world in these temporally asymmetric ways.

One common theme in Ismael's writing is that what prevents an Olympian conception of the world and our place in it is not "merely epistemic". In what follows we take up two related questions. First, what it would mean to say that these constraints are not merely epistemic, and second, how far these constraints get us in explaining why we experience the world in these ways.

We are inclined to think that these constraints may well be part of what is required fully to explain why things seem as they do to us, without any appeal to heavy-duty metaphysics; but in what follows we will suggest that they likely are not sufficient.

2. Prediction of the world from within and without

Let's begin by thinking about constraints on predicting the universe. Can we in principle predict the evolution of the universe, even given deterministic dynamical laws? A common error in thinking about this is that the answer is yes, all you need to do is to gather the information about how things are now, apply those deterministic laws, and project the rest of the universe's evolution.

Ismael (2016, this paper) rightly points out this is a mistake. Of course, if an agent *outside* of our space-time[2] had access to a time-like cross section of our four dimensional world—let's supposed this is a cross-section that is at right angles to the evolution of the cosmic background radiation to do some justice to one of the intuitions about what would make it a "time"—and the dynamical laws, then they could make such a prediction with the aid of a universe-sized computer. But no being *located in space-time* can do this.

Why not? Well, any being who has a space-time location has a backwards light-cone and a forwards light-cone. These are the regions of the universe that are casually connected to that being, intuitively causally prior in the first case and causally affected in the latter case. Then, just as our imaginary agent outside of the universe can take all the information to which they have access (the information about the laws and the time-like cross-section the universe) and make a prediction, so too our imaginary agent inside the universe would have to gather all the information they can, and likewise apply those laws and make a prediction.

But what information can the located being gather? Presumably only information that has had some causal impact on their senses or their measuring apparatus, and that restricts them to information in their backwards light-cone.

[2] If such a thing were possible.

Much of the universe, however, is not in either their backwards *or* forwards light cone. It's in the so-called "absolute elsewhere". It's located so that, transmitting causal influence at the speed of light, those signals cannot reach our imaginary agent's location. So, there is no way to gather *that* information at *that* location. Importantly though, that information can still have an effect: it can affect things in the agent's *future* light cone. So, the evolution of the universe from any particular location is affected by how things are in parts of the universe which cannot *in principle* be discerned by an agent at that location. So, prediction is in principle impossible for a located agent like us (and indeed presumably any possible located agent) even if the dynamical laws are deterministic.

The point was very nicely dramatized in 1966 By Larry Niven in his short story "At the Core." In this story the protagonist discovers (courtesy of alien faster than light travel, of course, so not in any way possible given what we take to be the laws of nature) that the galactic core has exploded, and a vast wave of radiation is heading towards earth. But from Earth's perspective the skies look normal, and will until such time as the killer wave reaches them, because the killer wave is travelling at the same speed as any sub-luminal information about it. So given all the information about the universe that is accessible consistent with the laws of nature, Earth would not have been able to predict its imminent destruction.

This is of course all true, and much neglected by philosophers who talk of predicting the future by finding out what is true "now" and applying the laws of nature. But if "now" is a hyperplane crossing into the absolute elsewhere this is impossible, and if it isn't, there isn't enough information in the "now" to perform the prediction. In either case, then, it cannot be done: it's nomologically impossible for an agent at a spacetime location.

What, though, does this tell us about why we experience the future as open? ? Ismael insists that these facts give us a kind of "freedom", and that this freedom is not *merely epistemic*. But what counts as "merely epistemic"? The picture of the world which underwrites this account of freedom is one on which there is (or would be if determinism were true) information which entails the future evolution of the universe, but on which that information is in principle, not merely in practice, inaccessible to embedded agents.

Whether all this adds up to physics setting us free, as claimed in the title of Ismael's 2016 book, is another question. But the matter at hand is whether it explains our sense of (for example) the openness of the future – or any of the other phenomenological features that are our explanatory targets. We doubt that by itself it does. Because if all it takes to be more than "merely epistemic" is for the predictability to be nomologically impossible, as Ismael seems to think, we think that its being so is explanatorily trumped by the fact that it was already *in practice impossible*. That is to say, even if such prediction were not nomologically impossible for embedded agents, it would still be in practice impossible and, it seems to us, learning that it is in fact also in principle impossible does not seem to add anything to the explanation of why things seem to us as they do.

To be sure, we *cannot* collect the microphysical data from a time-like slice of the universe and then predict the future. But it's also true that we cannot collect even the *local* microphysical data and predict via the dynamical laws either. We just don't have the sense, data storing or computational resources to do this. We certainly could not do anything like that during the period of our evolutionary history when

our phenomenologies evolved. Nor for that matter have we ever really been in a position to gather complete sets of available *macroscopic* data and apply *macroscopic* generalisations to provide detailed predictions, even if macroscopic generations were good enough to allow it. All this is to say that we cannot in practice come close to being able to collect all the relevant microphysical data in our backwards light cones: data to which we have in principle access. And this is sufficient to explain why we can't predict the future. What extra explanatory benefit do we get by adding to this the claim that the thing we have never been able to do in practice, for good practical reasons, is also in principle impossible? It's true, but surely if our inability to explain the future is what explains our experience of the future as open, then the fact that we are practically unable to predict the future is sufficient; and indeed, surely what explains the ways things seem to us, the phenomenological adaptions as we might put it, is that as a matter of fact prediction is fraught. That we could not overcome these practical difficulties is likely true, and interesting, but more needs to be said to show that this is explanatory.

3. Predictability of the world via coupled systems

Perhaps the most suggestive point Ismael makes is that systems embedded in a deterministic world will also be unable to predict what is going on insofar as they also are acting in it, and that somehow combining this with facts about the contingent local anisotropy of thermodynamics will get us much closer to an explanation of the target. Her thought is that it's the *interaction* of these two sources of unpredictability which is critical to the explanation.

The second aspect of this account is well known, and applies to the world as a whole, regardless of whether there are agents in it. If David Albert (2000) is right, we should help ourselves to the Past Hypothesis: the hypothesis that in the local past there is a very low entropy boundary condition such as for example the Big Bang. Given that there is such a low entropy local boundary condition, statistical mechanics alone will tell us that entropy will tend to increase away from that local boundary. Since we are located close to that boundary, entropy will decrease in one direction (away from the boundary) and increase in the other (towards the boundary) thus explaining at least some appearance of temporal asymmetries around here. For present purposes let's suppose that we have reason to accept the Past Hypothesis. Then we know that, around here, the world will contain many more records of how things *were* macroscopically than it does record-equivalents[3] of how the world *will be*.

Even granting all this, though, it is not obvious how, or indeed if, this explains the relevant phenomenological asymmetries. Perhaps consideration of the neuroscientific consequences of this being so will shed light on these asymmetries, but as far as we know this is yet to happen. As far as we can see, however, the mere presence of the asymmetries of records does not, in itself, explain why things seem to us to be the way they do. To see this, notice that what the Past Hypothesis gets us, is a difference in *degree* between past and future records, not a difference in *kind*. There are more records of the past, than record-equivalents of the future. But it is consistent with this being so, that there be creatures who have a phenomenology of remembering both the past *and* the future, although these creatures remember *much more* of the

[3] In case you think it's analytic that a record is past-directed.

past than they do the future. This, however, is not how things seem to us. It does not merely seem to us, phenomenologically, as though there is some difference in degree in the ways we know about the past versus the future, but rather, a difference in kind. So, we think that insofar as this asymmetry of records plays an explanatory role here, more will need to be said.

To be clear, this is no objection to Ismael's proposal. The explanation she offers combines this asymmetry of records with the view that agents are unable to predict their own local environment, and in particular to predict *themselves*.

She starts by making the point that any agent embodied in a world is subject to paradoxes of self reference. *Any* agent at all is so subject of course. Even an *unembodied* agent won't be able to give a good answer to the question of who predicts the acts of the agent who predicts the actions of all the agents who do not predict themselves! But perhaps there is no answer to *that* question. Ismael makes the point early on in her paper that there are questions to which there is an answer, that are similarly fraught for an embodied agent. Ask an AI "Is the answer you are about to display on your screen to this question "No"? Self reference in the question make this simple physical fact something that the agent can't predict, because if the answer displayed is "NO" it should answer "YES" and vice versa. Of course, you might think that there is nothing special about this from the point of view of prediction: no-one can provide an answer a question like "if Fred is the AI that services all the AIs that don't service themselves, who services Fred?". But the real point here is that for a device that is acting in the world, there is a kind of interference effect which makes prediction problematic. What will happen in the world depends on what the device is going to do, and so that requires the device to predict its own states. But a device which controls what it is going to do is unable stably to predict its own states, insofar as the process of control requires it to regard alternatives to the prediction as under its own control: it can 'bilk' any prediction of its own states. But that means its prediction of the world is unstable, insofar as what it does affects the world.

For what it's worth we think that the idea that a complete system can't predict its own states reliably is right, and is also right on independent, but perhaps not unrelated, grounds. This independent argument relies on the idea that a finite state machine can't predict its own states except by running itself – i.e. its own operation is the quickest prediction of what it will do.[4] The thought is this. In order to predict what a machine will do in micro detail, you need a representation of the machine to be added to the machine, plus processor space to act on that representation. But this adds to the machine, and insofar as the output of this prediction might affect the rest of the machine, that also needs also to be represented and predicted. But this adds further to the machine, which in turn requires further addition to the machine. This of course iterates infinitely, so no finite state machine will be able to predict itself.

If we are finite state machines, as surely we are, then we will not - on Ismael's or our grounds - be able to predict ourselves. And if predicting ourselves is required to predict the world, then there will be an interesting kind of inability for a being that acts in the world to predict the world. So here it is location as an agent rather than location in space-time which is the source of the limitation, and this is perhaps the most promising place to look for an explanation of some of the phenomenological asymmetries. We

[4] See Braddon-Mitchell, (ms)

have records of our past actions, or at least some of them. Retrodiction isn't bilkable, but the future seems open because we *cannot predict* what we will do more effectively than by doing it - just waiting until we decide then acting. And this, the thought goes, explains why it seems to us that the past is not open in the way the future is.

Well, maybe. Again, we think this is a part, perhaps an important part, of the picture. But more needs to be done to turn it into an account of the relevant asymmetries.

The first thing to note is that even if it's true on either Ismael's or our grounds that a machine cannot predict itself in complete detail, that doesn't rule out rough and ready prediction. Each of us often has a pretty good idea of what we'll decide. And while *complete* prediction at the micro level may be in principle impossible for the sorts of reasons just articulated, it is also practically impossible for the sorts of reasons we mentioned earlier. But then, so too is *complete retrodiction* at the micro level. Does this undermine the idea that this asymmetry might explain our phenomenology?. Not entirely, because you might point out that while we can both predict, and retrodict, in a rough and ready way, the asymmetry of records means that retrodiction is somewhat, even a lot, easier than prediction. But, again, this will be an asymmetry of degree, not of kind. Whether this is enough by itself to explain the phenomenological asymmetries is an open question - and until it's included in serious psychology, we won't have an answer.

But there is more to say here. You might think, as Ismael and Price (1996, 2007) do, that while we can predict ourselves in a kind of rough and ready way, we cannot do so *whilst taking the agential stance* on ourselves. And this is critical for explaining our phenomenology of the future unfolding, and the past being fixed.

The agential stance, however, is a metaphor, and one which needs to be unpicked. What's the idea? Perhaps we need to think about how rough and ready predication is possible, if the arguments above for failure of self-prediction in general are correct.

Consider our argument above: the problem there was that if we had an internal system predicting ourselves, insofar as its predictions would affect the output of the whole system, that system would need to be represented and predicted, and this would infinitely iterate.

This consideration seems to apply for predictions at any level, not merely the neurological level. Insofar as internal system A is predicting another internal system B, and A can potentially intervene in B, system A needs a prediction of itself in order to predict B. But then A needs more resources than is needed merely for predicting B, and so on.

How, then, can the commonplace that we are able to make pretty good predictions of our own decisions in the same way that we make predictions about the choices and actions of others be true? The answer, we think, lies in the phrase "A can potentially intervene in B". There is nothing to stop us from having internal systems that monitor and predict the behaviour of other internal systems *provided those predictions don't impact on the predicted process*.

So, consider an example. Pippa is considering whether to accept an academic position in Hong Kong. She currently works in Michigan. She gathers together a huge pile of information about the relative merits of the jobs and deliberates for a while. Her friend Janet looks on wondering "I don't know why she bothers will all this research and deliberation, there's no way she's going to Hong Kong, she'll decide to stay here". Janet has no problem making this prediction: the outcome of *Janet's* prediction (at least if she keeps it to herself) has no impact on Pippa. But now notice that Pippa,

taking a break from deliberation, might have the same thought "why am I putting all this work into decision making, I'm almost certain to stay in Michigan. Various well-known features of my psychology make it so" After her cup of tea, she gets back to deliberating, and eventually decides to stay in Michigan. She (and Janet) were right, and for the same reasons.

So, what's going on here, and why is it compatible with the idea that no system can predict itself? Well, when we look at Pippa we see that she has two uncoupled internal systems. To put it crudely, there's a deliberating module and a module which aims to guess on more general grounds what the deliberations module will do. And they are uncoupled: what Pippa thinks is likely or even certain of herself has no impact on the outcome of the module. That's why there is no issue here.

So, what is the 'deliberative stance'? To say that one is not able to predict one's own decisions from the deliberative stance is just to say that the deliberation module, or the sum of the deliberative processes, cannot predict itself. One's predictions of the outcomes of deliberation are predictions performed by *other* modules or processes, distinct from and uncoupled to, the deliberation module. This is not to say that these other modules or processes are not involved in deliberation over other matters, or even *related* matters (I'd be crazy not to renew my visa despite the huge cost, because likely I'll decide to stay, despite what I seem to think I ought to do while deliberating), it is just to say that they are not involved in the deliberation whose output they are aiming to predict.

So how much does this fact about deliberation and the agentive stance contribute to the overall sense of a fixed past, and a future coming into being as we decide and act?

If stances were global psychological facts, such that you either had the deliberative stance on yourself or you didn't, it might be extremely persuasive. If that were true, then either you would be an agent, engaged in acting in the world, regarding yourself as the author of your decisions, or you would not. If you did not take the deliberative stance on yourself you'd be some kind of wanton, simply watching your own internal processes, with a crippling sense of un-control and maybe a pathological phenomenology of time.

The agential perspective would, on this view, be the result of an integrated self, where every module or process affects every other one, and so no reliable predictions of your own behaviour would be possible. So, your own path through time would seem like a path in which the past is fixed, and the future unfolds partly as you make it. A lot more would be needed to spell that out of course. But it's far from clear that agents *are* integrated in this way. Perhaps they are not. If they are not, then while it's true that a *module* that is deliberating over P cannot predict whether it will choose P, it can still be that some other module can predict the output of that very module. If that is how things are, then taking the agential perspective is not something that whole persons or minds do: it something that modules do, or aspects of minds do. Then we are just left with the observation that a deliberative process cannot itself predict what it'll decide other than by deciding. But other process can, so long as they don't interfere, and sadly our prediction about our own behaviour frequently do not interfere with our behaviours in any direction.

If this is how things are, would this explain why we have the phenomenology that we do? We think this is not obvious. If this is how things are we can look into ourselves and see our deliberations as processes that can be predicted as much as can any natural system. Of course, the deliberative process themselves can't do that, only other

processes which don't interfere with the first. That's an important caveat; but it's not clear that it's enough to constitute a story about why we seem to be oriented in time the way many people say we seem to be. If we can predict the outcomes of our own deliberations, albeit only in *this* manner, then it's not obvious why we would feel as though the past is closed, and the future open, in quite the way that we do.

Perhaps, then, the upshot of these remarks is that as suggestive and powerful as these considerations are, there is much work to be done to use them as a basis for a complete explanation of our experience as temporal beings. Perhaps, also, they may shed light on the diversity of our experience. For instance, perhaps agents differ in regards to how unified they are, and perhaps this plays some role in people's having somewhat different temporal phenomenologies.[5] The current authors are excited by the prospect of examining whether there is any kind of 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, or strong intuitive differences in opinion about the metaphysics of time. Perhaps the natural A-theorist, for example, is the agent that can't take a distanced perspective on their own phenomenology.

At any rate, we think that while what Ismael offers us is the beginnings of an explanation of our experience of ourselves in time, it is just the beginning, and significantly more work in psychology and neuroscience will be required to complete that story.

Disclosure Statement

No potential conflict of interest was reported by the author(s).

References

Albert, D. (2000). Time and Chance. Harvard University Press.
Bardon, A. *A Brief History of the Philosophy of Time*. (Oxford: Oxford University Press, 2013).
Braddon-Mitchell, D. (2013). "Against the illusion theory of temporal phenomenology." *CAPE Studies in Applied Ethics*, ii: 211–233.
Braddon-Mitchell, D. (ms) "Self-prediction in finite state machines."
Dainton, B. (2011). "Time, Passage and Immediate Experience." in Craig Callender (ed.), *Oxford Handbook of Philosophy of Time*. (Oxford: Oxford University Press, 2011), pp 382-419.
Dainton, B. (2012). Time and Temporal Experience. In: A. Bardon (ed.). The Future of the Philosophy of Time. New York: Routledge.
Deng, N. (2013). "Our Experience of Passage on the B-theory", *Erkenntnis* lxxviii(4): 713-726.
Hodroj, B., Latham, A.J., Lee-Tory, J, and Miller, K (forthcoming). "Alethic Openness and the Growing Block Theory of Time" *The Philosophical Quarterly*
Hoerl, C. (2014), "Do we (seem to) perceive passage?" *Philosophical Explorations*, xvii: 188–202.
Ismael, J. (2016) *How Physics Makes Us Free* OUP.
Latham, A. J. Norton, J. and Miller, K. (2020). "An Empirical Investigation of Purported Passage Phenomenology". *The Journal of Philosophy*. 117(7): 353-386.
Lathan, A. J., and Miller, K. (ms) "Why do people represent time as dynamical? An investigation of temporal dynamism and the open future." Accessed at https://philpapers.org/rec/LATWDP
Le Poidevin, R. (2007). *The images of time: an essay on temporal representation*, Oxford: Oxford University Press.

[5] We already know that non-philosophers report quite different levels of agreement to claims about whether time seems to them to pass (see Latham, Miller and Norton 2020). We also know that people report having different views about whether the future is open in various ways. See Hodroj, Latham, Lee-Tory, and Miller (forthcoming) and Latham and Miller (ms).

Miller, K., Holcombe, A., & Latham, A. J. (2018). Temporal phenomenology: phenomenological illusion versus cognitive error. *Synthese*, DOI: 10.1007/s11229-018-1730-y

Miller, K. (2019). "Does it really seem as though time passes?" in *The Illusions of Time: Philosophical and Psychological Essays on Timing and Time Perception*, ed V Artsila, A. Bardon, S. Power, and A. Vatakis, pp 17-33. Palgrave McMillan https://doi.org/10.1007/978-3-030-22048-8_2

Niven, L. (1966). "At the Core." *In Words of If* . Quinn Publications.

Paul, L. A. (2010). "Temporal Experience." *Journal of Philosophy* cvii (7): 333–359.

Price, H. (1996). *Time's arrow & Archimedes' point: new directions for the physics of time*. New York: Oxford University Press.

Price, H. (2007). Causal perspectivalism In H. Price & R. Corry (Eds.), *Causation, physics, and the constitution of reality: Russell's republic revisited*. Oxford: Clarendon Press.

Prosser, S. (2016). *Experiencing Time*. OUP

Sattig, T. (2019) "The sense of temporal flow: a higher order account." *Philosophical Studies* 176:3041–3059