INTERFACE FOCUS

royalsocietypublishing.org/journal/rsfs

Research

Cite this article: Ismael J. 2023 Reflections on the asymmetry of causation. *Interface Focus* **13**: 20220081.
https://doi.org/10.1098/rsfs.2022.0081

Received: 7 December 2022
Accepted: 8 March 2023

One contribution of 15 to a theme issue 'Making and breaking symmetries in mind and life'.

Subject Areas:
biophysics, mathematical physics, systems biology

Keywords:
causation, interventionism, causal asymmetry, statistical mechanics, thermodynamic gradient, temporal asymmetry

Author for correspondence:
Jenann Ismael
e-mail: jismael1@jhu.edu

THE ROYAL SOCIETY
PUBLISHING

Reflections on the asymmetry of causation

Jenann Ismael

Department of Philosophy, Johns Hopkins University, Baltimore, MD 21218-2625, USA

JI, 0000-0002-3854-2898

The most immediately salient asymmetry in our experience of the world is the asymmetry of causation. In the last few decades, two developments have shed new light on the asymmetry of causation: clarity in the foundations of statistical mechanics, and the development of the interventionist conception of causation. In this paper, we ask what is the status of the causal arrow, assuming a thermodynamic gradient and the interventionist account of causation? We find that there is an objective asymmetry rooted in the thermodynamic gradient that underwrites the causal asymmetry: along a thermodynamic gradient, interventionist causal pathways—scaffolded intervention-supporting probabilistic relationships between variables—will propagate influence into the future, but not into the past. The reason is that the present macrostate of the world, in the presence of a low entropy boundary condition, will screen off probabilistic correlations to the past. The asymmetry, however, emerges only under the macroscopic coarse-graining and that raises the question of whether the arrow is simply an artefact of the macroscopic lenses through which we see the world. The question is sharpened and an answer proposed.

1. Introduction

Before Newton, it was largely assumed that causation was a paradigm example of an objective worldly relation. Early science treated it as the fundamental ordering relation of the world. The discussion of causation in physics began with a 1903 paper by Bertrand Russell in which he observed that physics in its modern, mathematical form did not incorporate anything at the fundamental level that looked like causation. He argued that the successor to causal notions were time-symmetric laws of temporal evolution and proposed that we eliminate talk of causation in scientific contexts in favour of those [1]. Russell's position turned out to be infeasible for a cluster of reasons, but it did raise the question of where the asymmetry comes from given the time symmetry of the underlying dynamics. The asymmetry of causation is arguably the most important of a cluster of temporal asymmetries that characterize our experience of the world. In the last few decades, two developments have shed light on the asymmetry of causation: (i) the introduction of the interventionist analysis of causation and (ii) increasing clarity in the foundations of statistical mechanics. Recent work taking advantage of those developments, however, raised questions about whether (and in what sense) the asymmetry might be perspectival (see [2–7] and especially [8]).

This paper is an attempt to attain clarity on the issue. We ask what precisely is the status of the causal arrow, assuming a thermodynamic gradient and the interventionist account of causation? We find that there is a perfectly objective asymmetry rooted in the thermodynamic gradient that underwrites the causal asymmetry. Along a thermodynamic gradient, interventionist causal pathways—i.e. scaffolded intervention-supporting probabilistic relationships between variables—will propagate influence into the future, but not into the past. The reason (we will see) is that the present macrostate of the world will screen off probabilistic correlations to the past, in the presence of a low entropy boundary condition. That observation connects the causal asymmetry directly to the asymmetric boundary condition that generates the thermodynamic arrow.

The asymmetry, however, emerges only under the macroscopic coarse-graining. We assess whether (and in what sense) that relativization makes the causal arrow perspectival.

2. Causation

Why was Russell not right? What does causal language add to mere descriptions of patterns of correlations? And why do we need that information? Consider the example of smoking: there is a correlation between smoking and getting cancer. There is also a correlation between having smoker's breath and getting cancer. Quitting smoking, however, is a good way to reduce the risk of cancer but treating bad breath is not. Why? Because smoking is a *cause of* cancer, having bad breath is not. Capturing that difference—the difference between information and influence, i.e. between correlations that allow one event to carry information about another and correlations that allow one event to (be used to) influence another—is what causal information is for. What causal language adds to patterns of correlations is information about the routes by which influence travels. We need that information so that we know how to intervene in the orderly flow of events strategically. For creatures like us—creatures who do not merely observe but intervene, and who want to know how their interventions will affect what happens—causal knowledge is indispensable. Nancy Cartwright pointed this out in an influential response to Russell in 1979 [9]. In philosophy, interest turned to analyses of causal notions. People tried to reduce causal information to special patterns of correlation; counterfactual accounts became popular but got bogged down in trying to produce a semantics for counterfactuals. There was some confusion about whether we were trying to capture people's pre-theoretic intuitions about causation or give an account that one could relate to the parts of science that are actively involved in the causal search. The issue remained unresolved.

It re-emerged in connection with the foundations of thermodynamics and attention to the physical underpinnings of temporal asymmetries, but without a clear (and precise) analysis of causal notions, it was difficult for the discussion to get traction.

Everything changed with the introduction of interventionism. Interventionism, which was introduced independently by Pearl [10] and Glymour, Spirtes and Scheines [11] and developed philosophically by Woodward [13],[1] provides an elegant formal framework that captures the idea that underlies the search for causes across all sciences. The heart of interventionism is the idea that A is a cause of B if and only if it there are circumstances in which it is possible to manipulate B by intervening on A.[2] An intervention is an idealized, unconfounded experimental manipulation intended to cut ties between the manipulated variable and other past causes, so that any remaining correlations can be assigned to the manipulated variable. Interventionism provides a precise understanding of what causal knowledge added to probabilities, the kinds of data that are used to establish causal relationships and the patterns of reasoning those relationships support, grounded in experimental practice. It also provides a formal calculus, analogous to the probability calculus, for representing and reasoning about causal relationships.[3]

Even though scientific knowledge is naturally conveyed and conceived in causal terms, there had long been a bias in terms of presenting it formally in statistical or probabilistic terms. The idea was that causal talk was heuristic and the factual content was conveyed by statistics. Interventionism removed any basis for that bias.

3. Statistical mechanics

There has been a question about the sources of temporal asymmetry in the world since the time symmetry of the classical dynamical laws raised questions about the source of the asymmetry captured in the second law of thermodynamics. A century and a half of work on the foundations of statistical mechanics showed how to recover the emergent dynamical asymmetries even though the underlying laws are temporally symmetric.

The Boltzmannian account of statistical mechanics given by David Albert in *Time and chance* [18] is an exceptionally clear and careful conceptual articulation of those foundations.[4] The account assumes three things:

(1) The classical dynamical laws: these are just the familiar Newtonian laws of motion.
(2) The statistical postulate: this is the central postulate of statistical mechanics; it is the standard Lebesgue measure over phase space that gives the probability that a system in a given macrostate is in one or another of the microstates compatible with that macrostate (or, more precisely, is in some subvolume of the volume of the phase space associated with that macrostate).
(3) The past hypothesis: this is a boundary condition in the distant past, most often framed as the hypothesis that the universe was once in a state of very low entropy.

These works together to produce thermodynamic generalizations into both the future and the past as follows. The dynamical laws delimit a set of physically possible worlds. The statistical postulate imposes a probability distribution over those worlds that heavily favours worlds on entropy-increasing trajectories.[5] The past hypothesis eliminates all worlds except those that were in a very low entropy state at some time in the distant past. The result is a history for our world that is overwhelmingly likely to be increasing in entropy. Of the three principles, (1) and (2) are symmetric in time. The Newtonian mechanical laws are symmetric under temporal reflection: for any physically possible trajectory of a closed system that leads from A to B, the reversal of that trajectory (obtained by reversing the positions and momenta) is also physically possible. The probability distribution is time independent. The asymmetry is embodied entirely in (3) which imposes an asymmetric boundary condition at only one end of time.

There are disputes about various elements of Albert's account. I will rely on it here but if it is not your preferred account of the foundations of statistical mechanics, you can substitute your own. Any account of the foundations of statistical mechanics that is successful at generating thermodynamic generalizations will underwrite the asymmetry of records. It is useful to see how it works on Albert's account, but whatever logic you use to derive the asymmetry of records from your own postulates, reversing that logic should allow you to derive the causal asymmetry in the form described below.

4. Asymmetries of knowledge

The project of trying to characterize temporal asymmetries in our experience and ground them in the thermodynamic gradient goes back to Reichenbach. It passes through Paul Horwich and Huw Price and gets developed with Albert.[6] There are two asymmetries: one epistemic and one practical. The epistemic asymmetry is about trying to capture the sense in which we know more about the past than the future. The practical asymmetry is about trying to capture the sense in which we can affect the future but not the past.

Albert's account of the epistemic asymmetry proceeds as follows: we start with an agent that has perceptual access to the present surveyable macrostate of its environment and then we use statistical mechanics to see what the agent can infer from that information.[7] So we take the microcanonical probability distribution over that macrostate and evolve it forward and backward, conditionalizing on the low entropy past.[8] The resulting probability distribution will leave the agent with a lot of specific detailed information about the macroscopic past, but very little about the future [18,23]. The reason is that conditionalizing the present macrostate of the world on a past hypothesis will allow the agent to infer (with overwhelmingly high probability) that all of the semi-ordered states of approximately isolated subsystems in its environment are evolving *from* even more ordered states, and that will turn them effectively into records of their past. So, for example, consider a half-mixed cup of coffee or a footprint in the sand, or indeed, any semi-ordered macrostate of an approximately adiabatically isolated subsystem of the environment. If we take the macrostate, apply the probability postulate and evolve it forward, we can infer with overwhelming probability that it is heading towards a state of higher entropy (cream will be more mixed, footprints will fade). If we evolve the same distribution backward *without* conditionalizing on the past hypothesis, since the dynamical laws are time-symmetric, we get the same thing in that direction: it is overwhelmingly likely to be coming *from* a state of higher entropy. But if we conditionalize on a lower entropy state in the past, everything is different. We consider only those trajectories coming from states of lower entropy: so we can infer that half-mixed cream was less mixed and footprints more pronounced. And if we trace those trajectories back to the even that initially produced the ordered state, we find that half-mixed cream is a record of the introduction of a drop, a footprint is a record of a footstep and a scar on a tree is a record of a past injury.

What about the practical asymmetry, i.e. the idea that affects the future but not the past? The contours follow what we said about records. We take the same ingredients and ask how changes in the world of the kind that agents like us can bring about propagate into the past and future.[9] So, an agent asks: what will happen if I walk across a sandy beach, dig a hole, bury some nuts or manipulate a local feature of the environment? What will happen, that is to say, if I do some work to create an ordered state in the environment? To answer the question we take the present surveyable macrostate of the world, apply the probability postulate to get a distribution over microstates compatible with that, evolve it forward using the dynamical laws and backward, conditionalizing on the past hypothesis. What we find is that agents that are perceptually coupled to the world macroscopically will see the effects run into the future, leaving the past untouched. So, for example, if I walk across a sandy beach, there will be footprints until the tide washes them away. If I dig a ditch or build a house, I am working in more durable materials, but doing the same thing: creating an ordered state in the environment that will take some time to decay. The asymmetry of records and the asymmetry of effects are flip sides of one another. What you think of as the effects of your actions are just the future records of their occurrence. An intelligent agent uses her interventions strategically to create records of her actions that she will encounter later.

Looking at the logic of this clarifies why interventions leave the past untouched: conditionalizing on the past hypothesis fixes every feature of the past of which there is a macroscopic record. Any feature of the past of which there is even the possibility of a macroscopic record is going to be insensitive to local macroscopic interventions in the here and now. That means that once we conditionalize the present macrostate on the past hypothesis, there will not (with overwhelming probability) be any reliable, emergent probabilistic correlations to past events.[10] Putting these pieces together, then, will provide (the schema of) an explanation of why creatures like us, who view the world through macroscopic lenses, see the effects of our own actions as running into the future.

So far, we have not explicitly said anything about causation. We assumed the laws, probability postulate and past hypothesis, and we have said what kind of information an agent with perceptual access to the macrostate of her environment has available to her about the past and future and what sorts of expectations she could form about the results of her own actions. A quick route to the asymmetry of causation would say that that is all there is to the causal asymmetry, but that leaves too many questions unanswered.

We want to know what is there on the side of the agent: what concepts does she bring to bear on her experience? How are causal concepts embedded in the network of concepts that she brings? What inferential connections do they bear on other concepts and how are they wrapped up in her practical and epistemic exchanges with the world? We also want to know what there is on the side of the world: what is the external infrastructure in the environment that supports the application of those concepts? Since the agent is herself part of the physical world, we can make the whole unit—agent + environment, or better, agent *embedded in* environment—the unit of physical analysis. The interventionist account gives us an account of causation that does not make any essential reference to the perspective of the agent. We use the framework to get a clear diagnosis of whether there is an asymmetry that can be identified prior to the introduction of agents.

5. Interventionist causation

According to interventionism, given a network of variables and a set of constraints, a variable A is causally related to B (relative to that network and under those constraints) just in case interventions on A (surgical changes in the value of A) affect (or affect the probability of) B. Below we have a network of variables. We want to know if A is a cause of B, so we intervene on A and see if that affects (the value or the probability of values of) B (figure 1).

There is no restriction to macrovariables, so the A's and B's here can be anything. The framework is temporally

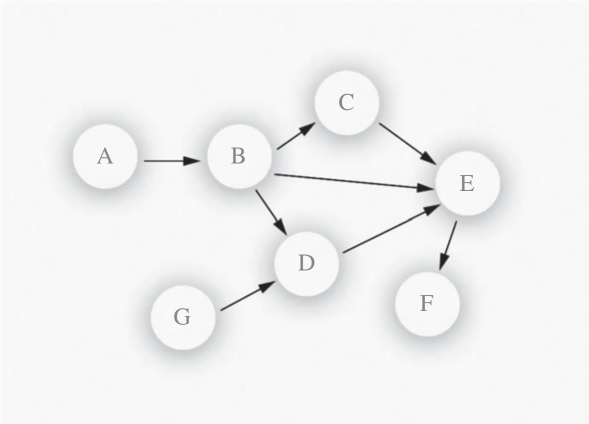

Figure 1. Causal graph.

symmetric. Normally, we are looking for causal effects in the future, so we define interventions in a temporally asymmetric way (interventions are surgical changes in the value of a variable that cut ties with other potential causes in the past) and look for causal effects on temporally downstream variables. If we want to know whether taking aspirin relieves headaches, we ordinarily fix the environmental constraints that define the experimental set up together with the broad surveyable macrostate of the world and the past hypothesis. There is nothing in the interventionist framework, however, that forces us to think that way. We could just as easily take a network of variables, choose a variable A, surgically cut ties with temporally upstream variables, and ask whether interventions on A will affect (the probability of) values of variables in the temporal past.

Pearl discusses the asymmetry in his 2000 book [28]. In his view, the directionality comes entirely from the choice of what to treat as exogenous and what to treat as endogenous. One carves off a piece of the world, and thinks of oneself as reaching in, setting the values of certain variables and observing the effects on the others. The variables whose values are set by processes outside the network are exogenous and the others are endogenous. This is what he says:

> [The] choice of [endogenous and exogenous variables] creates asymmetry in the way we look at things, and it is this asymmetry that permits us to talk about 'outside intervention', hence, causality and cause-effect directionality

We tend to choose exogenous variables earlier than endogenous variables because questions about how later states vary with differences in early ones have a special importance for purposes of guiding action. Formally, however, there is no problem choosing exogenous variables later than endogenous ones. We can raise questions about the effects of variation in future states on the past as surely as we can about the effects of past states on the future. Such questions are logically well behaved.

> 'The lesson', he says 'Is that it is the way we carve up the universe that determines the directionality we associate with cause and effect. Such carving is tacitly assumed in every scientific investigation. In artificial intelligence it was called circumscription, by McCarthy. In economics, circumscription amounts to deciding which variables are deemed endogenous and which ones exogenous, IN the model or EXTERNAL to the model' [28, p. 420].

On this view, there is no intrinsic direction of dependence in the relations among events. The world itself has a modal substructure (given by time-symmetric laws) that furnishes a basis for judgements about what would happen in hypothetical situations defined by a choice of exogenous and endogenous variables (given specified constraints). The direction of dependence is determined by the choice of exogenous and endogenous variables. Those are in their turn defined by the practical interests of the experimenter. We are temporally oriented agents, so we tend to hold the past fixed and look for causal effects downstream, when we are deliberating, but there is no intrinsic asymmetry in the relations among events themselves.[11]

That is not quite *incorrect*, but it is incomplete in a way that is misleading. Filling out the picture will give us a deeper and more explicit understanding of the relationship between the asymmetry of agency and the asymmetry of causation. A preview: it is going to turn out that there is not a local asymmetry in the relationships among events, but there is an asymmetry in the macroscopic *pattern* of events along a thermodynamic gradient that agency emerges to exploit. So the explanation goes from the existence of the macroscopic asymmetry to the emergence of agents. And the reason that we see through macroscopic lenses has to do with the way that asymmetry supports information-gathering and utilization.

Return to what was said earlier. We saw that the classical laws, past hypothesis and microcanonical probability distribution will underwrite causal relationships that run from past to future. What if we asked the time reversed question? So we fix the broad surveyable macrostate of the world. We reach into a network, get a hold of some variable, cut ties with variables in the future, and ask what happens to earlier variables if we wiggle later variables. Will there be any probabilistic effect on variables in its past? So long as we fix the low entropy past, the answer will be no. The present surveyable macrostate together with the low entropy past will fix every past event about which there is a macroscopic record and that will screen off any potential probabilistic effects running into the past.

From a statistical mechanical point of view what is going on is that, if we understand causal relationships interventionistically, the past hypothesis is among the constraints that we generically impose and that will be enough to secure a causal arrow as a matter of objective fact. The low entropy boundary condition in the temporal past is not *just* a 'constraint that we generically impose'; it is a fact about the world, something that is part of the fixed background against which we act. It is part of the unvarying scaffolding in our world that lets us anticipate the future effects of interventions and know things about the past. If the 'direction of causal influence' is the direction in which the probabilistic effects of local interventions propagate, then in a world with an entropic gradient, causal influence propagates into the future leaving the past untouched. The interventionist framework—because it is itself temporally symmetric—helps us locate (or make explicit) the source of the asymmetry.[12]

We can remove agents from the picture and use the interventionist analysis to bring into focus an asymmetry in the physical setting in which agents act. Along a thermodynamic gradient, causal relations—i.e. scaffolded intervention-supporting probabilistic relationships between variables—will not run from future to past for very general reasons: viz. because the present macrostate of the world will screen off probabilistic correlations to the past. Since there is no analogous boundary condition in the future, probabilistic influence will propagate freely in that direction.

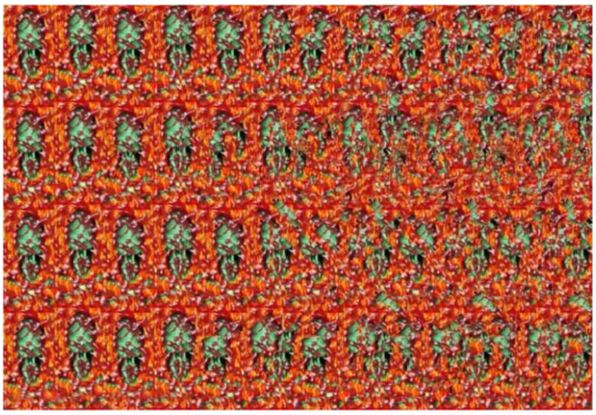

Figure 2. Magic eye picture.

6. The macroscopic coarse-graining

We have not entirely eliminated the shadow of the agent, however, because we are still talking about the macrostate. The notion of a macrostate is a little ambiguous; sometimes it is used in a generic sense to refer to any coarse-graining of the microstate and sometimes it is used to refer specifically to the coarse-graining imposed by thermodynamic variables. Up to this point, I've been using it in the specific sense without worrying about the ambiguity, but it will matter in what follows, so in this section, I will use lower-case for the generic sense and capitalize when I mean the coarse-graining imposed by thermodynamic variables. Of all the ways of carving the world into different coarse-grained states, is there anything that makes the thermodynamic coarse-graining special other than that it is the one that our senses happen to pick out? Even if there is an objective asymmetry that emerges under the Macroscopic coarse-graining, why are we talking about the Macrostate?

Let us suppose that we all agree that if our senses could see right down to the microscopic level, we would see no direction of determination in the laws. And let us suppose that we all agree that when we have the whole structure assembled—laws, thermodynamic gradient, agent coupled to the world's macrostructure, intervening on local variables—temporal asymmetries emerge from the agent's perspective that match our own. There remains the question of whether we see the effects of our actions propagating into the future but not the past simply *because we happen to see through Macroscopic lenses*.

The way to press the question is to ask what other perspectives might be possible on the same world. We take a world that is just like ours at the microlevel: same microstate, same time-symmetric laws. We introduce an agent; the sensors of the agent are going to pick out and 'light up' certain coarse-grained variables. Everything else is pushed into the background as the invisible network behind the scenes that controls the dynamics of the variables we can see (figure 2).

By coarse-graining over different macrovariables, we can bring into relief different patterns. Choose any set of macrovariables and the microdynamics induces a dynamics over those variables (tells us how they evolve in time). We couple to the world via macrovariables that obey the second law of thermodynamics. That means that we see irreversible processes and the effects of our own interventions running into the future. But formally we can coarse-grain in different ways. Allowed free reign in a structure of any complexity, we can introduce agents that will 'see' many different patterns. Maybe there are ways of finding a bird or a fish, or the word 'Jesus' coming into view one letter at a time. Looking from the bottom up, there is nothing that announces the pattern that we see as having any special status. The question that one would like to ask is whether we could take the same world, introduce an agent and couple it to the world by making a choice of sensors and actuators that would reverse its temporal perspective. *Could we, that is, fix the microstate of our world just as it is and even find a way of coupling to it that would reverse the temporally oriented features of our experience?*

It seems like there ought to be a sharp answer to the question. As it stands, however, it is radically underspecified. How much leeway do we have in choosing coarse-grainings? Do we allow highly gerrymandered redistricting of phase space? Do we allow only space–time integrals of conserved quantities? And what sort of interventions do we allow? Do we allow Maxwell Demon style control of microscopic degrees of freedom or distributed high-level variables like inflation? What about the Hamiltonian of universe? And how do we motivate constraints? We do not want to choose constraints that match our own perspective because the point is to explore what alternative perspectives there might be. With complete free reign the answer should be trivially yes: we just take the ordinary thermodynamic coarse-graining, for any volume of phase space, take the set of trajectories coming out of that volume, and reverse them. Now choose some time interval (e.g. 100 years) and take the set of points of the reversed trajectories as defining a new volume of phase space. It is not going to look like a coarse-graining of the kind to which we are accustomed; it is going to be a highly fibrillated set of points, but we know by Louisville's theorem, that volume is preserved. Let us refine the question to get at something philosophically interesting. What do we really want to ask here?

The tendency in the philosophical literature, when thought experiments are introduced to separate what is in the world and what is an artefact of perspective, is to treat ourselves as transcendental subjects coming from the outside and coupling to a world that we are not a part of. That is a mistake here. We should not be thinking in terms of transcendental subjects tapping into the world from outside. We should be asking what ways of seeing the world there might be by *agents in the world* with sensors that pick up information via physical channels and actuators that manipulate local features of the environment; we should be thinking, that is to say, in concrete terms about what sorts of *embodied* causal perspectives there might be in a world like *ours*.[13] There is a whole cluster of physical questions that need to be nailed down before we have fully precise questions: minimally an agent should be a system with sensors and actuators; it should recognize the distinction between what it sees and what it does; it should be able to learn over time about the effects of its interventions and use that information to guide behaviour. An agent like this will have an *internal* arrow defined by the temporal direction in which it sees the effects of its actions as propagating. A recent, quite beautiful paper by Pete Evans, Gerard Millburn and Sally Shrapnel gives us exactly the kind of physical analysis we need to make the question precise and answer it. They introduce a minimal model of a causal agent and show that the internal arrow of any such agent is going to be aligned with the thermodynamic gradient. The simple reason is that the physics of such an agent is dissipative, and not for accidental reasons, but for reasons associated with the connections between energy and information which are at the heart of the issue here. Any physical system that functionally gathers and uses the information to guide behaviour is going to be using energy and is going to be subject to thermodynamic constraints.[14]

7. The heart of the matter

That is half of the story. It tells us why, if there are information-gathering and using systems, they are aligned with the thermodynamic gradient, but it does not yet tell us *why there are information-gathering and using systems*.

Here is how hydropower works. We build a dam so that when water passes through a sluice, it flows down a channel and turns a turbine (figure 3).[15]

The turbine is a subsystem of the world with an internal arrow of rotation. The internal arrow is aligned with the direction of the external current. Why? Because of physics. The physics of the device shows why there are turbines that run in the direction of the current and not counter to it. But what if we asked not 'why does the rotational arrow of turbines align with that of the external current?', but 'why are there turbines?'. The answer to that question is more revealing and to the point for our purposes here: *turbines exist precisely to exploit that gradient*.

We saw above that the temporal orientation of causal agents will be aligned with the thermodynamic gradient. What if we ask, not just 'what kinds of causal agents are physically implementable', but 'why are there causal agents?'. Why are there creatures with sensors and actuators that learn and what does that have to do with the thermodynamic gradient? What kinds of agential perspectives would arise naturally and indigenously in a world like ours? The answer is illuminating; agents emerge to exploit the opportunities for information utilization and control that are created by the thermodynamic gradient. It is not an accident that we coarse-grain and coarse-grain in the way that we do: our senses are designed to reveal the asymmetric, information-rich patterns give agency a grip.

Microscopically, there are local processes, happening in accord with unchanging laws. Macroscopically, irreversible processes—both locally in adiabatically isolated subsystems and at the global level—happen in the direction of increasing entropy. The low entropy boundary condition in the past makes it possible for information about the macroscopic past to accumulate in the form of records.[16] The macroscopic world where there is a thermodynamic gradient is littered with records that contain the imprint of its macroscopic history. A footprint in the sand, a scar on the trunk of a tree, a photograph, a series of letters on a piece of paper …; all of the semi-ordered approximately adiabatically isolated systems you see around you are evolving from states of even lower entropy and bear the imprint of their past.

The macroscopic development of those systems left to their own devices will follow its ordinary course, dissipating energy into the environment: footprints will wash away, ice will melt. But the information contained in the macroscopic environment (i.e. in the present state of systems evolving from lower to higher entropy) is available to *other* systems to use as a basis for their own behaviour. And evolution

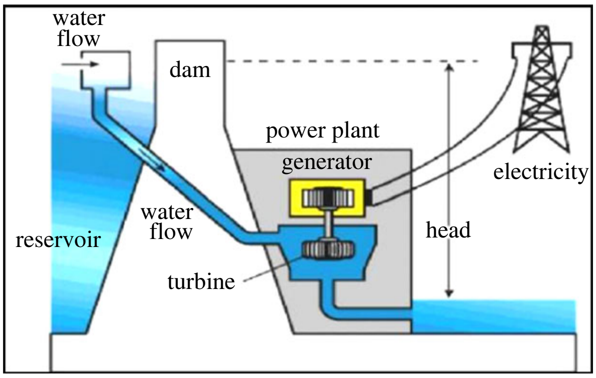

Figure 3. Water turbine in a hydropower plant.

populates the universe with systems that do that. A deer that can read the signs of a recent lion hunt or a fox that knows that a hole in the ground means a rodent is likely nearby will do better than one that does not. A creature that has some information about what has happened, has information about what is going to happen. Because there are macroscopic regularities that link what *has happened* to what *will happen* that are not screened off by the macroscopic present, there is a selective advantage to using the information in the environment.

Like the deer and the beaver, we rely uncritically on records for information about the past in our daily lives. We take it for granted that chalk marks on a blackboard, tracks in the snow, drawings on ancient cave walls or even craters on the moon are remnants that were once more ordered states. The foundations of thermodynamics aim to make explicit the physical facts that underwrite that reliance. If the universe did not start in a low entropy state, the local macroscopic environment would not carry any more information about the past than about the future. The things we think of as signs of predator and prey would have been much more likely to have fluctuated out of equilibrium by chance than to have evolved from a more ordered state. All of this is possible because the low entropy past makes information about the macroscopic past readily available. It is because of the low entropy past that facts about the past, although long gone, leave an imprint on the current macrostate of a system that can, in its turn, be used to regulate local processes. So even though the microprocesses are all local and Markov (probabilistically screen off their own past), we see higher-level processes that effectively build causal bridges between the past and the present by using the information-bearing properties of macrostates in the presence of a low entropy boundary condition in the past.

So, it is not an accidental fact about us that we have sense organs that track macroscopic information. Indeed, once we are given the macrostate of a thermodynamic system, conditionalizing on the microstate does not (typically, under the kinds of conditions that we find out in the wild) affect probabilities for the macroscopic future. That means that once one knows the macrostate of a system, one has typically extracted everything from the past that can be usefully parlayed into information about the future. What all of this suggests is that our perspective here is carefully crafted to reveal the information-rich macropatterns on which agency works. By filtering out the microscopic noise, your perspective is lightening the load of cognition (or doing some of the cognizing, depending on how you like to think of it).

So, if one looks from the bottom up, the macroscopic coarse-graining is not going to look any more or less special than the one that reveals the word of Jesus. It is not going to have any special *metaphysical* status. But if one is searching in the assembled architecture of the world for a stratum that will support information-gathering and utilization, the pattern that emerges under the macroscopic coarse-graining will stand forth as special indeed.[17]

Suppose that we were suddenly given eyes that could see the world in full microscopic detail. We can ask whether adding microscopic information would increase future macropredictability. In a deterministic setting, if we have complete information about the microscopic state of the universe at one time, we can predict everything that would happen at other times with certainty. But if we know anything *less* than everything, that power goes away. The microscopic laws entail that the future behaviour of an open subsystem of the universe is a product of its internal microstate together with exogenous influences. Generically, knowing the microstate of a system is not going to help us predict behaviour unless we also know all of the microscopic influences that might impinge on it from the outside: everything from the position of molecules in the air to neutrons raining down from the atmosphere. Anything short of complete information at the microlevel will not ensure the prediction of the future microstate. If we want to apply the microlaws to derive predictions, we either need full information or the ability to shield exogenous influences. That is a kind of control we can often arrange in the laboratory, but do not ordinarily find out in the wild so there is no clear selection pressure to enhance our perceptual perspective with microscopic information, and the thermodynamic cost of capturing and recording the information provides pressure against it. In practice, it is not the deterministic microscopic laws, but rather the emergent patterns that are robustly *indifferent* to microscopic underpinnings that make prediction possible. The most general of those are the laws of thermodynamics. Those apply universally and (virtually) without exception. In addition, there are all kinds of macroregularities about the typical behaviour of specialized subsystems: trees, frogs, people, or penguins. Those are what—in practice—make the world predictable.[18]

There are evolutionary models to help us understand the conditions under which behaviour regulated by sensitivity to various environmental cues would be selected for. Peter Godfrey Smith, for example, provides a formula for calculating the expected payoff of attending to various kinds of information in the environment for regulating behaviour [32]. He is using it to ask questions such as should sea moss always have spikes or should it have sensors that detect a chemical produced by sea slugs and produce spikes only when there is enough of that around? Under what circumstances will the flexible strategy (the strategy of having behaviour regulated by local environmental parameters) be better than the best inflexible one? Whether it is better to be flexible depends on whether the cue being used is reliable enough to overcome the difference between the expected importances of the two states of the world. The *importance* of a state of the world is defined as the difference between the payoff resulting from appropriate action in that state and the payoff resulting from an alternative (inappropriate) action in that state. The *expected importance* is defined as the importance multiplied by the probability of that state. If the organism is going to produce only a single behaviour, it is best to produce the behaviour suited to the state of the world with the higher expected importance. It is worth using a flexible strategy only if the cue associated with that strategy is reliable enough to overcome the asymmetry between the expected importances of the two states of the world.[19]

The model addresses the question: what are the properties of reliability that a cue must have before it is worth using to direct behaviour? We can apply it to gauge the expected payoff of using microscopic versus macroscopic information to guide behaviour in an environment like ours. What we are suggesting is that given macroscopic information, microscopic information would have little or no payoff under typical conditions.[20] All of this reinforces the lesson that the macroscopic state of the environment is where the information-rich patterns on which cognition works are. Microinformation does not help except under conditions that are not typically available in the wild. So the right response to the worry that the arrows that emerge under the macroscopic coarse-graining are just accidents of perspective is that there is nothing accidental about our perspective.[21]

8. Conclusion

We started with the concern that the temporally oriented features of our experience were 'mere' artefacts of perspective and what we have found is that our perspective is rather tightly integrated into the world and tailored to bring into relief the information-rich patterns on which cognition—and in particular, cognition, in the rich form that we exemplify it—works. Perception and cognition are evolved activities of embodied beings that cannot be disentangled from the underlying physics of the world in which they operate. They arise to exploit the opportunities physics provides.

Is this a form of perspectivalism? It is not glib perspectivalism. Glib perspectivalism about some feature of the world says that it is merely an artefact of the lenses through which we see it. So, for example, you might accuse someone of seeing the world through rose-coloured glasses, implying that the rosiness is entirely projected, i.e. that the rosiness is not in the world but in the glasses and they are misattributing it if they attribute it to the world. The contrast to glib perspectivalism is thoughtful perspectivalism, where one thinks, 'The world is full of good and bad. The person who sees the world through rose-coloured glasses is seeing something real, attending to the good in the world, allowing it to dominate their field of view'. The thoughtful perspectivalist is someone who sees the world through lenses that reveal real patterns.[22] Someone can see the world through the economist's glasses, attending to economic variables, or see the world through gendered lenses that bring gender relations into relief. They can see it through lenses that highlight information flow or power relations. In these cases, what they see is perfectly real. The lenses reveal something that is there in the world. It is an objective question which patterns are there and in some cases, for given purposes, which ones are worth seeing.

Data accessibility. This article has no additional data.

Conflict of interest declaration. I declare I have no competing interests.

Funding. This paper was based partly on research made possible by a grant from the Foundations Questions Institute and Fetzer Franklin Fun (FQXi grant number FQXi-RFP-IPW-1906).

Acknowledgements. I am grateful to David Albert and Carlo Rovelli for inspiration, comments and so much else. Many thanks also to two excellent anonymous reviewers who provided very valuable recommendations.

Endnotes

[1]See [10,11] for the formal framework and Woodward [12] for philosophical development, including discussion of the notion of intervention.

[2]The formal definition of an intervention remains a point of debate (see [13]). Interventionism does not provide the only causal notion that can be clearly defined and proves useful to science. Causal process notions [14,15] capture aspects of our causal intuitions that are not captured by the interventionist conception of cause. My view is that there is no single concept that captures all of those intuitions and that we should rather be pluralists about the collection precisely definable physical concepts that answer to different causal intuitions. The reason for focusing on the interventionist notion here is that causal processes are not intrinsically asymmetric. It is the interventionist pathways that are the target for analysis of the causal asymmetry.

[3]On the comparison between the probability calculus and interventionist formalism see [16,17].

[4]See [18]. For a recent and especially nuanced assessment that deals also with the quantum case, see [19].

[5]Increasing, here = non-decreasing.

[6]Albert's book [18] and then later in Albert [23]. See also [24,25].

[7]Details are given in Albert [18] and [23].

[8]Of course, no real agent actually has full information about the macrostate of the world so this characterizes the information that is in principle accessible to the agent. Substituting 'surveyed' for 'surveyable' will tell us what information an agent actually has.

[9]There is a bit of subtlety here about how to characterize events in the world that are rightly thought of as coming under agential control. Human movements themselves do not have an inherent grain: if you raise your arm, for example, you also raise all of the microparticles of which it is composed, produce electrical impulses that travel up the muscles, and induce microscopic neural changes in your brain. But control requires perceptual feedback. There are complex sensorimotor networks in the human body that loop in things like movements of our arms and legs, the motion of our head, the sound of our voice, without any more fine-grained voluntary control.

[10]Albert [18,23] offered a counterfactual account of causation and tried to derive the causal asymmetry from the thermodynamic gradient. Frisch [26] and Elga [27] proposed counterexamples involving special conditions engineered to secure counterfactual dependence of past events on present actions. Albert conceded that under the right conditions his account would generate counterfactual dependence of past events on present ones and hence there would be backward causation but argued that it was not the kind of dependence that could be used strategically to bring about past events. The backwards dependence in both examples (which trades on the fact that present events can provide information about the past) disappears when the event in question is an intervention because interventions break the kind of past probabilistic dependence on which they depend. Interventionist causal pathways are scaffolded intervention-supporting probabilistic relationships. The absence these kinds of relationships running into the past is the mirror of the reason that there are no current records of the future.

[11]That is what I argued effectively in (2016) [29]; everything depends on what you hold fixed and what you allow to vary.

[12]If the low entropy past is not imposed as a constraint, the past effects of present interventions (cutting ties with upstream variables and asking about effects of interventions on the past) precisely mirror the future effects of present interventions. Once we impose the low entropy boundary condition, the asymmetry emerges with probabilistic effects running only into the future.

[13]The notion of a perspective has always had a double life. One can treat perspective formally and ask whether some domain permits one to introduce a formal reference frame of a certain kind and how some feature of interest depends on the reference frame. Or one can treat a perspective as something that is physically realized. In space–time physics, for example, sometimes people mean a formally defined frame of spatial reference and when they talk about symmetry transformations, like boosts and rotations—they mean a purely formal operation. The idea is that one could leave the world entirely alone and make a mathematical transformation to our frame of reference that leaves the laws (or the model or the system under discussion) intact. Other times, they mean a physically embodied frame of reference that is part of the system that is being modelled, e.g. a moving ship, a person in an elevator, and symmetry transformations are physical changes in the location or state of motion of the frame. In that setting, the distinction was often pointed out and largely benign. It becomes important here because we are talking about the alignment of a system's internal dynamics with that of its environment where asymmetries are in play.

[14]http://philsci-archive.pitt.edu/18844/, https://arxiv.org/abs/2009.04121. Their paper is wide-ranging and argued in detail. As soon as one is thinking in these terms, connections between energy and information are going to become important and it is going to begin to look like (something approaching) law that any physically implementable system that uses and processes information is going to be aligned with the thermodynamic gradient. See [30,31].

[15]An anomymous reviewer points out correctly that a more apt analogy might be furnished by wind turbines or solar panels which exploit flows that arise naturally. That observation brings out the central suggestion here, which is that the world provides affordances that agents emerge to exploit. Evolution produces agents whose senses are sensitive to macroscopic variables because that is where the exploitable information-bearing patterns lie.

[16]I am reverting to the ordinary convention of not capitalizing 'macrostate' even though using it specifically in the remainder of the paper to refer to the thermodynamic coarse-graining.

[17]The existence of emergent macroregularities robust under the kinds of noisy conditions we find out in the wild is already highly non-generic, but what happens to these laws along a thermodynamic gradient is what supports information-gathering and utilization with a long horizon. In denying that that the macroscopic coarse-graining has a special metaphysical status, I mean anything that requires ontological distinction not captured by these dynamical considerations.

[18]People often say things like (assuming determinism) if you know the exact brain state of someone making a difficult decision you would be able to predict their decision. But of course, that is nonsense. In principle, as a matter of microphysical law, anything can make a difference to your behaviour from the particular velocity with which a particular molecule of air strikes the surface of their skin, to a tiny dust particle that lands on the back of their neck. The response is often made that these things are not likely to make a difference, but in saying that, one relies on macrogeneralizations about what makes a difference to the decisions of creatures like us. The same goes for the behaviour of trees and traffic lights.

[19]This model is closely related to a model of information gathering used in Bayesian decision theory which describes the benefits of experimentation. It is supposed that an agent must make a decision about how to act in a given situation and that she has the option of undertaking a cost-free experiment. The possible outcomes of the experiment are (for her) probabilistically related to the states of the world that determine the success or failure of her action. A cue associated with a flexible strategy that is inferior to an inflexible strategy in the biological model corresponds to a zero-value experiment in the Bayesian model.

[20]'Environment like ours' means a classical world of similar complexity, along a thermodynamic gradient. For another approach to a closely related conclusion, see [33].

[21]Rovelli [34] conjectures that in a world as complex as ours there ought to be a coarse graining that reverses this whole suite of features. That raises the question of whether there could be subsystems of our world, coupled to it by macroproperties which define a notion of entropy whose gradient is reversed relative to thermodynamic entropy. This is a different and much more radical challenge than I have answered here. The possibility of formally coarse-graining in a way that reverses the gradient leaves open that there are constraints on coarse-graining that rule this out as a concrete possibility, but I leave this challenge as an open question.

[22]The term is used deliberately to evoke 'real patterns' [35].

References

1. Russell B. 1913 On the notion of cause. *Proc. Aristot. Soc.* **13**, 1–26.
2. Price H. 1991 Agency and probabilistic causality. *Br. J. Philos. Sci.* **42**, 157–176. (doi:10.1093/bjps/42.2.157)
3. Price H. 1992 Agency and causal asymmetry. *Mind* **101**, 501–520.
4. Price H. 2007 Causal perspectivalism. *In Causation, physics, and the constitution of reality* (eds H Price, R Corry), pp. 250–292. Oxford, UK: Clarendon Press.
5. Woodward J. 2003 *Making things happen: a theory of causal explanation*. Oxford, UK: Oxford University Press.
6. Woodward J. 2007 Causation with a human face. In *Causation, physics, and the constitution of reality* (eds H Price, R Corry), pp. 66–105. Oxford, UK: Clarendon Press.
7. Woodward J. 2015 Methodology, ontology and interventionism. *Synthese* **192**, 3577–3599. (doi:10.1007/s11229-014-0479-1)
8. Woodward J. 2021 *Causation with a human face*. Oxford, UK: Oxford University Press.
9. Cartwright N. 1979 Causal laws and effective strategies. *Nous* **13**, 419–438.
10. Pearl J. 2009 *Causality*. New York, NY: Cambridge University Press.
11. Spirtes P, Glymour C, Scheines R. 1993 *Causation, prediction, and search. Lecture notes in statistics, 81*. New York, NY: Spring-Verlag.
12. Woodward J. 2003 *Making things happen: a theory of causal explanation*. Oxford, UK: Oxford University Press.
13. Woodward J. 2001 Causation and manipulability. In *The Stanford encyclopedia of philosophy (winter 2016 edition)* (ed. EN Zalta). Stanford, CA: Stanford University.
14. Dowe P. 1992 Wesley Salmon's process theory of causality and the conserved quantity theory. *Philos. Sci.* **59**, 195–216. (doi:10.1086/289662)
15. Anon. 2008 Causal processes. In *The Stanford encyclopedia of philosophy (fall 2008 edition)* (ed. EN Zalta). Stanford, CA: Stanford University.
16. Pearl J. 2001 Bayesianism and causality, or, why I am only a half-Bayesian. In *Foundations of Bayesianism. Applied logic series*, vol. 24 (eds D Corfield, J Williamson). Dordrecht, The Netherlands: Springer.
17. Geffner H, Dechter R, Halpern JY (eds). 2022 *Probabilistic and causal inference: the works of Judea Pearl*, vol. 36, 1st edn. New York, NY: Association for Computing Machinery.
18. Albert D. 2000 *Time and chance*. Cambridge, MA: Harvard University Press.
19. Wallace D. 2021 Probability and irreversibility in modern statistical mechanics: classical and quantum. *arXiv*, 2104.11223. (doi:10.48550/arXiv.2104.11223)
20. Horwich P. 1987 *Asymmetries in time*. Cambridge, MA: MIT Press.
21. Price H. 1996 *Time's arrow and Archimedes' point*. Oxford, UK: OUP.
22. Price H, Corry R (eds) 2007 *Causation, physics, and the constitution of reality: Russell's republic revisited*. Oxford, UK: Oxford University Press.
23. Albert D. 2015 *After physics*. Cambridge, MA: Harvard University Press.
24. Hoerl C, McCormack T, Fernandes A. 2022 *Temporal asymmetries in philosophy and psychology*. Oxford, UK: Oxford University Press.
25. Rovelli C. 2019 *The order of time*. New York, NY: Riverhead Books.
26. Frisch M. 2010 Does a low-entropy constraint prevent us from influencing the past? In *Time, chance and reduction* (eds G Ernst, A Hüttemann), pp. 13–33. Cambridge, UK: Cambridge University Press.
27. Elga A. 2001 Statistical mechanics and the asymmetry of counterfactual dependence. *Philos. Sci.* **68**, S313–S324.
28. Pearl J. 2009 *Causality: models, reasoning, and inference*. New York, NY: Cambridge University Press.
29. Ismael J. 2016 How do causes depend on us? The many faces of perspectivism. *Synthese* **193**, 245–267.
30. Bennett CH. 1982 The thermodynamics of computation—a review. *Int. J. Theor. Phys.* **21**, 905–940.
31. Maroney O. 2009 Information processing and thermodynamic entropy. In *The Stanford encyclopedia of philosophy (fall 2009 edition)* (ed. EN Zalta). Stanford, CA: Stanford University.
32. Godfrey-Smith P. 1996 *Complexity and the function of mind in nature*. Cambridge, UK: Cambridge University Press.
33. Shalizi C, Crutchfield J. 2001 Computational mechanics: pattern and prediction, structure and simplicity. *J. Stat. Phys.* **104**, 816–879.
34. Rovelli C. Is time's arrow perspectival? *arXiv*, 1505.01125. (doi:10.48550/arXiv.1505.01125)
35. Dennett DC. 1991 Real patterns. *J. Philos.* **88**, 27–51.