Causal Fictionalism

Antony Eagle

2023-05-22

Contents

A few years ago now I argued that causation, while not a relation of fundamental physics, was nevertheless pragmatically indispensable (Eagle 2007a). A number of other philosophers made similar arguments around the same time (Price 2007; Menzies 2007), and there are many precedents in the literature. Here I want to revisit these arguments with the benefit of hindsight. I don't in essence disagree with what I said back then, but I think what I say here both significantly clarifies my earlier discussion and advances things to some degree.

In the first part of the present chapter I start with the role of causal models in the human sciences (§1), and attempt to explain why it is not possible to straightforwardly ground such models in fundamental physics (§2). This suggests that further constraints, going beyond physics, are needed to legitimate such models (§3). These supplementary constraints could be reified, but that would seem to conflict with the completeness of physics (§4). A response is to emphasise the

practical role of causal talk (§5), and I suggest that a fictionalist approach might be worth exploring. After clarifying fictionalism as a general approach (§6), in §§7, 8, 9 I try to carry out in some detail the project of clarifying what a fictionalist attitude to causation would involve.

1 Causation in the Human Sciences

Ramsey observed that

> from the situation when we are deliberating seems to arise the general difference of cause and effect. We are then engaged not on disinterested knowledge or classification ... but on tracing the different consequences of our possible actions. (Ramsey [1929] 1990, 158)

Cartwright (1979) makes the related point that causes are vital in drawing a distinction between effective and ineffective strategies for attaining our desired ends. When we are deciding which actions to perform, we ought generally to pursue actions which genuinely promote our goals. The standard decision theoretic approach to deliberation says we ought to pursue those actions with the greatest expected value to us (Peterson 2017). Here 'expectation' is a probabilistic notion; Cartwright's central point is that the probabilities involved need to capture causal information (1979, 431).

Some strategies can be seen to be ineffective on purely statistical grounds. The probability of being susceptible to mind reading, conditional on wearing a tin foil hat, is very low. But that is due to the low unconditional probability of mind reading, which is probabilistically independent of tin-foil-hat-wearing. This is an ineffective strategy because it is probabilistically irrelevant. By contrast, there was a strong non-accidental association between the construction of airstrips and the delivery of clothing, medicine, and other technologically advanced goods throughout Melanesia during World War II. But the 'cargo cult' strategy of constructing mock airstrips didn't further the participants goals, because the association was the symptom of a common cause.[1] Building airstrips was an ineffective strategy because it involved probabilistic dependence without causal relevance. An effective strategy would have probabilistic dependence that was backed by causal relevance: for example, the strong association between the deployment of malaria nets and the decreased incidence of childhood malaria is backed by a plausible causal mechanism (Levitz et al. 2018).

The human sciences – the social sciences, health and medical science – are focussed not on 'disinterested knowledge', but on providing a basis for action to improve people's lives (given some antecedent conception of wellbeing). It is not

[1] The real anthropological story, of course, is more complex (Lindstrom 1993).

surprising, then, that even cursory glance at the literatures of the human sciences suggests they are replete with invocations of causation. Attributions of blame and responsibility, evaluations for the efficacy of interventions, and tests to discriminate causal from non-causal associations are basic parts of the standard conceptual toolkit in economics (Hoover 2008), public policy (Cartwright and Hardie 2012), clinical medicine (Williamson 2019; Stovitz and Shrier 2019), and public health (Hill 1965; Lucas and McMichael 2005), among many others. The key results – and the principal reasons for our interest in these fields – are causal claims: claims that a certain drug, or public health initiative, will lower the incidence of negative outcomes from illness or disease (i.e., cause them to be less frequent); or that adoption of some policy will promote a desirable social or economic outcome. Across these fields we see a trans-disciplinary deployment of techniques of statistical causal inference to secure such results. These techniques aim to deliver exactly what Cartwright argued is needed to discriminate effective strategies: a way of identifying non-causal ('spurious') associations (Granger 1969; Simon 1977; Suppes 1970; Pearl 2000, 42–57).

However, these techniques are often strikingly empiricist, in the ancient sense (Dawes 2017, sec. 3.3): theoretical mechanisms are discounted in favour of observed associations among the values of random variables, preferably ascertained by a systematic review of randomised controlled trials (Howick et al. 2011). The evidence consists of observable values of random variables, a theoretically-laden classification of 'raw experience' into usable data. These random variables partition ('coarse-grain') the space of possible outcomes into classes whose members all agree on the value of that variable. The variables whose values can be aggregated into statistical data cannot be arbitrary functions from outcomes – to be *data*, different outcomes must be observably distinct from one another, so that the value of the variable in a given state can be discerned. For these variables are to be set exogeneously in the causal model, rather than having their values fixed endogenously within the model, and while purely theoretical variables could in principle play this role, in practice the values of exogenous variables are determined by statistical observation. This refined empirical data is then used to generate a causal model depicting a pattern of relations between the random variables (they may postulate unobservable hidden variables too). Generally, there are many models consistent with the data, and we prune the models by eliminating any that don't capture all and only the observed statistical dependencies between variables. This process can terminate in a unique model only in the case where there are no hidden variables. These models give us causal relations that accord, more or less, with key platitudes about causation that might be taken to constitute a 'folk theory' (Norton 2003, sec. 2.5): that causes temporally precede their effects, that causation is acyclic, that causes make their effects happen.

In practice, only some of the causally relevant factors are identified as causes,

others being relegated to *background* (or *boundary*) conditions. The distinction here is grounded in the possibility of intervention, the manipulation of a causal variable in such a way as to isolate its impact on some effect variable. A genuine cause is such that manipulating it can trigger a corresponding difference in value in the effect variable, in a way that manipulating a non-cause cannot (Eagle 2007a, 167). The technique of randomised experiments (Fisher 1935), deriving ultimately from Mill's 'method of difference' (Mill [1874] 1974, bks. III, ch. VIII, §§2–3), is a widespread application of this idea, where the randomization of subjects to treatment and control groups aims to ensure match of causally relevant properties, and hence to enable any treatment effect to be observed in the aggregate impact on the effect variable. Without the possibility of intervention, this methodology is unrealizable. It may be unrealizable in practice anyway, but causal inference techniques rely on the mathematical possibility of intervention so as to 'sever' a causal variable from potential confounders (Pearl 2000, 157–58, 348–51). There is no reason why we can't define a variable whose values are grounded in conditions across a vast swathe of space and time, but obviously such a variable could not have its values arbitrarily fixed by local experimental interventions. Moreover, such variables block the application of standard inference techniques, because their presence in the model potentially undermine the ability to isolate causes from their effects by surgery on a structural equation model. For this reason, the grounds for the values of causal variables are typically localised in space and time.

Even in those sciences where the randomized controlled trial is deprecated in favour of *models*, such as engineeering or geology, local variables play a key role. The construction of a causal model in those fields is not theory-neutral, as RCTs promise to be, but rests on a particular conception of causal mechanisms as *connections* between variables. The mechanism is a 'black box'; to understand the mechanism is to open the box and examine the detailed linkages between parts of the process within. In this light, a mechanism can be contained, being itself localised and having localisable parts. As Woodward puts it,

> Mechanisms consist of parts, the behavior of which conforms to generalizations that are invariant under interventions, and which are modular in the sense that it is possible in principle to change the behavior of one part independently of the others. (Jim Woodward 2002, S366; see also Cartwright 1999)

From all this, a picture emerges: the causes involved in the human sciences and many non-fundamental branches of natural science are manipulable, spatiotemporally localised event types; and an ideal causal model in such a science should, while reflecting observed statistical associations, also give us grounds for deciding which interventions are effective in making desired outcomes happen.

2 Effective Strategies and Physical Laws

As many have observed (Norton 2003; Field 2003; Latham 1987; Russell 1913; Cartwright 1979; Frisch 2014; Price 2007), the construction of causal models in the human and special sciences is strikingly different from what goes on in the development of our most fundamental theories in physics. This is evident in both method and product. In method: because of the central role of theory in physical prediction and explanation, the principal direct aim is to discover experimental grounds for the acceptance of theories. Protocols like RCTs, which aim to establish causal conclusions without the intermediating role of models or theories, are thus much less relevant to the epistemic goals of physics. Mechanical models are more theoretically freighted, but even there significant isolating assumptions are required – most notably, robustness to variation in boundary conditions and localized encapsulation of the mechanism. These assumptions are absent in fundamental physics, which tends to have global ambitions.

These observations about method shouldn't be pressed too hard. Causal models, like those constructed via the graph-theoretic approach to causal structure (Pearl 2000; Spirtes, Glymour, and Scheines 2000), are clearly theories about unobserved aspects behind the statistical data. Even more obviously, mechanical models in geology, engineering, etc., make explicit and extensive use of global mathematised physical theory (though udner idealizing 'small world' assumptions). Maybe, despite not making any notable use of causal inference algorithms and going far beyond mechanistic reasoning, fundamental physics is also producing theories that are causal, through and through. It would be surprising if this is so, given how important a role substantive causal assumptions are in causal inference – as Cartwright (1989, ch. 2) has it, 'No causes in, no causes out' – and how little evidence in physics of any such assumptions being made. Still, perhaps physics uses a different means to arrive at the same causal end.

Accordingly, the main ground for thinking that causal structure is of less importance in physics lies in the content of the theories produced by physics. The theories of classical mechanics, relativity, and quantum mechanics don't involve *localised causal variables* in any obvious way. The don't involve any inference to dependency from coarse-grained associations between the occupants of assorted chunks of space and time. Those theories, quite different in detail, can be nevertheless all be understood very abstractly as involving a complex, perhaps many-dimensional, space (a spacetime manifold, or maybe a higher-order 'configuration space' in the case of some views of quantum mechanics), over which space are given geometric fields characterising the metric structure of the space, and material fields[2] characterising its occupants and the values of various phys-

[2] I am intentionally abstract here, attempting to capture the stress-energy tensor of GR and the quantum wave function under a single description.

ical quantities (Earman 1986, 32:24; Maudlin 2021; Wallace and Timpson 2010; Ney 2015; Albert 2013). Different models of these theories involve different specifications of the values of these fields across the fundamental space; that which all models of a given theory agree on fixes the laws of a given theory. A theory is thus a family of constraints on the compossible global distributions of field values across spacetime (or configuration space), 'all at once' as it were. These theories don't in any obvious way involve the production of the field values over time. Causation involves making things happen, and nothing in these models is a good candidate making relation.[3]

One obvious reply is that we ought not to look at the theory presented as a 'timeless block'. To do so misses out any *dynamics* in the theory, and doubtless the dynamics are candidate causal relations. In standard formulations, the geometrical structure of the underlying space of a theory will permit the specification of dynamical principles that govern how the material field qualities realised at some region are linked to those at nearby regions. For example, in a geometry which permits objective simultaneity relations, there will be instantaneous properties instantiated at times which correspond to rates of change in physical properties over time. The dynamics specifies how these instantaneous rates are linked to the instantiation of the property of which they are rates. For example, velocity is a rate of change of position, and a dynamics will specify how material velocities at some time are linked to material positions at nearby times. The dynamics, presented as a family of differential equations, give rise to the same timeless block model in terms of the global pattern of property instantiation. But the dynamics generates such a block from the bottom up, constraining how the states of different regions can be patched together to form a nomologically coherent whole.

With our interest in causation, the relevance of the dynamics derives from a proposal to understand dynamical laws as causal laws: the local state at some region causing the states that surround it. But despite an etymological affinity, dynamical laws don't fit the profile of properly causal models established in §1. For one, the constraints on stitching states together don't generally have any temporal orientation: a local state constrains its preceding neighbourhood no less than its succeeding neighbourhood. These are 'laws of association' (Cartwright 1979, 419), where the contents of a given region sufficient to fix the contents of other regions will fix its whole surroundings, not just its future surroundings. To get causes here, we need to stipulate that only precedent regions can count as causes.

More importantly, generically these dynamics do not give rise to dependence relations between localised states. Fix on some local state as a putative effect.

[3] Sometimes people talk of relations in such theories using causal language; for example, sometimes the lightcone structure in special relativity is called its 'causal structure' (Curiel 2021). But on closer inspection this turns out to be 'causal' only in the sense that if spacetime points are spacelike separated (outside the lightcone), that is sufficient for them not to be causally connectible of causal relations between them. No causal connections need be instantiated.

What generates that local state is enough of its neighbourhood to fix its character, given the dynamics. But this will generally not be itself a localised region. Take the case of relativity, which at least imposes some requirement of locality (through having an upper bound on physical velocity). Any spacelike surface intersecting the past light cone at a point p will fix the character of p; any less will not be enough information for the dynamics to work with to establish what p is like. Such a surface is not typically very localised, and will generally encompass a very large region. Moreover any such surface will manage to fix the character of the effect. So, dynamically speaking, any arbitrary slice through the past light cone is a cause; all such slices are equally causes; and any proper subregion of such a slice is not enough to be a cause.

If we want an effective strategy for manipulating the contents of a given spacetime point p, and we want to treat the dynamical laws as causal laws describing the operations of our strategy, then the targets of our interventions must be vast regions intersecting every trajectory of potential causal relevance. These include not only what we'd intuitively regard as potential causes, points along those trajectories which might be the paths of genuine causal (mechanical) processes leading to p – we must also specify the *absence* of certain events at any points on trajectories which could interfere with those processes. Such a specification will be completely *undiscriminating*, because every point in the past light cone is potentially occupied by an interfering event. This yields a notion of cause that is not the kind of thing that could be effectively manipulated; these 'causes' are too many, various, and far-flung. The dynamical laws tells us which events we need to specify in order to jointly fix the occurrence of the effect, but these events are not causes, because no humanly effective strategy can exploit or control them (Field 2003, 439). Likewise: take a discriminating recipe, that classifies some (but not all) localised spacetime regions as causes. We want causes that 'make the difference'; but intervene on such a region, holding fixed all the other causes, and the dynamical laws will not give any determinate fix on any localised potential effect, because we need also to fix the status of any potential interferers. Dynamical laws can help in identifying effective local interventions only if those local interventions are physically *isolated*, so that there is no potential influence from outside the region of the local cause (Elga 2007). But of course 'being isolated' is not an intrinsic property of a region; to establish that the dynamics can count a localised region as a cause, it needs to be established somehow that the contents of its surroundings are non-interfering. The dynamics alone won't establish this, because they are compatible with models in which local regions are not isolated.

3 Finding Causes in a Physical World

The upshot, I think, is this. Causal relations govern local interventions that effectively (but without guarantee) produce localised outcomes. The models of fundamental physics, even when represented dynamically, don't involve any obvious role for such relations, nor any obvious candidates to be, or surrogate for, them. This leaves us with three main options: elimination, reduction, and supplementation.

Elimination was Russell's preferred option:[4]

> the word 'cause' is so inextricably bound up with misleading associations as to make its complete extrusion from the philosophical vocabulary desirable. (Russell 1913, 1)
>
> [causal laws] though useful in daily life and in the infancy of a science, tend to be displaced by quite different laws as soon as a science is successful. The law of gravitation will illustrate what occurs in any advanced science. In the motions of mutually gravitating bodies, there is nothing that can be called a cause, and nothing that can be called an effect; there is merely a formula. (Russell 1913, 13–14)

Perhaps if our interest lay purely in describing physical reality, Russell's proposal would have some merit. But we ought not 'displace' our use of causal laws to identify effective strategies, in favour of the austere equations of fundamental physics. Even if it were practically possible to do so, many of our effective strategies aren't deemed such by the fundamental laws. The use of mosquito nets is an effective strategy to decrease the prevalence of malaria. This high-level truth won't fall out of the physics unaided; indeed, the pattern of correlations supporting the effectiveness of the strategy won't fall out of physics unaided. (Thus eliminativists who retreat to pure physics can't opt for 'diet causation', i.e., counterfactually robust correlations.) Certainly if we model an individual mosquito net and an individual mosquito as an isolated system we may be able to predict that the net will block any trajectory of the mosquito that begins on one side of the net and ends on the other. But this relies crucially on tacitly suppressing any possible interferers. Once we embed the mosquito net in the wider world, there will be many physically possible scenarios in which the net is rendered permeable to mosquitoes by some interfering process from outside the narrowly isolated system we were considering. Of course such scenarios are unlikely and not to be taken seriously in deliberation. But that judgment of unlikelihood, and the decision to set aside such scenarios when deliberating, isn't motivated by the physics alone.

[4] See also van't Hoff (2022, sec. 2), who argues that causation must be a discriminating relation, and that 'the insurmountable problem of selection' entails that causation is a 'relation which cannot be instantiated'.

Quite generally, suppose we were to try and capture effective interventions in physics by looking at all models in which some local event is realized in some localised region C – the kind of region we might conceivably intervene on in action. Turn the crank on the global dynamics in all of these models, and see what comes out in some target local region of interest E. We would get, in the general case, nothing especially interesting, because the nomological influence of the regions outside C would swamp the influence of C itself. (Here again: unless C is a significant chunk of the past light cone of points in E, the contents of C can be trumped by the contents of the region of the light cone outside C.) To get something out that shows the character of C to fix the character of E, in the absence of knowing the system to be isolated, we're going to have to have some sort of probability distribution over the possible contents of the rest of the initial value surface outside of C (Elga 2007, 118). Such a distribution will allow us to say that, with high probability, intervening on C will fix what happens at E – assuming we assign higher probability to those physical possibilities where no interferers emerge from outside C than to those possibilities containing one or more interferers. Even setting aside the difficulties in constructing a suitable probability function for use in causally-informed decision making – issues I return to below in §7 – nothing in the physics itself will give us such a probability distribution (Norton 2003, 9–10). It might come from our background knowledge, where we just weight the actual world and those worlds similar to it more heavily than other physical possibilities that are more dissimilar to actuality. Or it might come from a general epistemic preference for simpler hypotheses.[5] Neither way does this weighting come from the physics itself, which assigns no special nomological significance to actuality or to any credence distribution over physical possibilities.

This argument that we cannot satisfactorily provide surrogates for effective strategies in the models of pure physics is *a fortiori* an argument against *reduction* of causal laws to physical laws – for reducible causal laws would be perfect surrogates for themselves, assuming again that causal relations are central to the distinction between effective and ineffective strategies. So if we need causal relations to appeal to in deliberation and action, we're going to need to accept that they go beyond the patterns in pure physics. Accordingly, we're going to need to *supplement* physics with additional structure in order to see how it can support causal models.

Obviously we could supplement the physics with enough background information to uniquely specify the actual course of events, perhaps a global initial condition if the physics is deterministic. That would avoid any difficulty with interferers, for physics supplemented with this much information entails the occurence

[5] The possibility in which there are no interferers outside C is plausibly simpler than the worlds in which there are interferers, so a general Occamist preference for simpler hypothesis might make us weight those possibilities more highly, either in our rational credences, or in some *a priori* logical probability that weights simplicity strongly (Solomonoff 1964).

of everything that actually happened. With this sort of information, however, there is no role for causation; even the idea of an effective strategy is redundant, given deliberation is otiose under such conditions. Prediction on the basis of some event C ought to be guided by the outcomes in some class of physical models in which C obtains (and the differences across a matched population of models otherwise alike in which C does not obtain). The whole class undermines prediction, because almost anything is possible (interferers, the effect coming to pass via some other potential pathway) if we allow arbitrary variation in the non-C-grounding parts of the model. The narrowest class, the model corresponding to actuality, undermines causation because there is no sense to be made of marking C as a cause as opposed to any other condition accompanying a given outcome. So we need information that narrows the class of models in a way that shows the presence and absence of C to reliably indicate a certain effect. What sort of information will do this?

As above, we can add a probability distribution over the possible contents of the environment of a purported cause. This is *modal* information about what is likely possible for the surroundings of our cause. Another source of modal information that plays a similar role is counterfactual information, and the significance many accounts of causation give to counterfactuals reflects this (Lewis 1973; Hitchcock 2001; James Woodward 2003; Paul and Hall 2013). Counterfactuals concerning what would have been the case had A obtained impose some sort of structure over families of physical models. They divide all A-compatible models into those compatible with what would/might have been the case if it had been that A, and those that, while they are compatible with A, are not consistent with what might have been the case if A.[6] Indeed, even bare modal facts about what is possible for a given system will serve to supplement the physics, as long as they are facts about a species of possibility more restricted than nomological or physical possibility. For example, facts about what a given thing *can* or *has the ability* to do.

That causation requires supplemental facts that are not evident in models of fundamental physics also connects with observations made previously about the role of assumptions of isolation in order that the dynamics might support relations of local determination. When considering a given local region C, such supplemental information will enable us to 'prune' the space of physically possible models that include C. We can eliminate those models in which C's surroundings are too improbable, or its surroundings depart too gratuitously from actuality, or are not surroundings that are realistically possible. That is: this additional information will serve to specify a quite narrow range of permissible assumptions

[6] This allows us to set aside certain A-compatible possibilities given the laws, for the purposes of deliberation. Of course setting a possibility aside, and assigning it a significance proportional to its probability, are not the same thing – but for many practical purposes they will yield the same deliberative verdicts.

about what the background surroundings of a given potential cause are like, and hold those fixed in the class of models while we 'toggle' C and note the effects of doing so. In the good case, such information will reassure us that 'differences in distant matters of fact are *unlikely* to make a difference' (Elga 2007, 110, my emphasis). And when distant matters of fact are unlikely to make a difference, we can for all practical purposes treat a given system as isolated from its surroundings. Sometimes, of course, the background information tells us that the system isn't isolated, but its environment 'exerts a uniform influence throughout' (Demarest and Hicks 2021, 616). If so, the system is 'quasi-isolated', and the probabilistic and counterfactual information tells us not that the causal region suffices to fix the character of an given effect by itself, but that it manages to make the difference between the effect having one character versus another, relative to that fixed background. When a system can be treated as isolated or quasi-isolated, the dynamical laws and our causal models can happily fit together.

Often, of course, we don't have the slightest idea how to represent the variables of interest in our causal models in the vocabulary of fundamental physics. We might be able to come up with a causal model for the motion of a rock (Elga's example) that derives from the dynamics and an assumption of isolation; we aren't going to be able to do the same for epidemiology or agronomy. But the same assumptions of (quasi-)isolation continue to be important, perhaps even more so. In those cases we typically only have statistical data about macroscopic variables of interest, and no microphysical basis. To discriminate causal from non-causal associations we need some background information assuring us that any correlations that are not screened off by some other variable are indicative of causation. We also need some assurance that the causal surroundings permit genuine causes to appear as associations in the statistical data. (So they are not masked or interfered with – Hesslow (1976).) Both sorts of background information takes the form of assumptions that other possible explanations of the statistical data are excluded. And on what grounds do we exclude these other explanations? Typically, if they are too improbable (i.e., a given correlation is statistically significant for some stringent level of significance), or if invoking them would amount to a gratuitous departure from what we think would happen in similar cases (if we need to suppose some sort of 'cosmic conspiracy' in the surroundings to mask the role of a given variable).

Cartwright goes further, arguing that nothing less than information about causal factors will help us discriminate causal and non-causal associations: we need causal information in, to get causes out (Cartwright 1979, 423). Obviously an appeal to a causally homogenous background will provide modal information that performs the role played in the discussion above by probabilistic and counterfactual information, assuring us that our system is quasi-isolated from other potential influences. Cartwright's invocation of causal information is, in effect,

an assumption that the system is quasi-isolated. Then a cause is something that plays a counterfactual or probabilistic difference-making role against a fixed background of other causes.[7]

Let me sum up. Whether the supplementing information is causal or not, the consensus seems to be that causation cannot be located using the resources of physics alone. Information about causal dependencies provides information constraining *what might happen* more tightly than information about the *local* physical state does, even given the background physical laws. (Though not as tightly as the *global* prior physical state might, especially under determinism.) That tighter constraint is what makes such information useful for deliberation. (I give more details on *how* causal information plays a role in deliberation in §7.) To try and deliberate using just the resources of physics either requires an appeal to the voluminous body of potentially physically relevant information, which tightly constrains judgements about the outcomes of interventions, at the cost of being both cognitively unwieldy and recommending infeasibly global interventions; or appeals just to local information that provides little constraint on what a local intervention will entail. Moreover, identifying the *kind* of supplemental information that is needed seems to involve ideological resources that go beyond those of the physics: background information sufficient to fix causes isn't physically distinctive, and plays no independently motivated role in the physics (unlike, say, the global background state or the past light cone).

This applies likewise to the information about probabilities and counterfactuals that play a role in grounding the distinction between effective and ineffective strategies: those too cannot be located in the resources of pure physics. Probabilistic or counterfactual dependence also supplements physical law to enhance prediction in the presence of only local physical information. Hence those who agree with Russell that causation is to be eliminated, but who think it can be adequately replaced in deliberation by talk of counterfactual dependence, are equally committed on my view to non-reductionism about effective strategies.[8]

7 Appeals to modal and counterfactual information may also involve causation, though perhaps in a less overt way. For example, not just any counterfactual information helps us delineate the background against which a purported cause makes a difference to the data. Not just any truth that would have still obtained had P failed to obtain is part of the background against which P is to be evaluated as a potential cause, because if we allow backtracking counterfactuals ('had P not occurred, then it would have to have been that P'' [so as to ensure it is still the case that E]) then we end up treating background events as causally relevant. Arguably, drawing this distinction presupposes some causal information, as non-backtracking counterfactuals are 'already assumed to be, in a broad sense, causal' (Hitchcock 2007, 58). Even the notion of isolation itself might be a causal notion: 'A system is isolated (in our sense) just in case all of the features that have causal influence on the suitably specified "output" state of the system go through the "input" states' (Demarest and Hicks 2021, sec. 3.5).

8 Dorr (2016, sec. 7) is no causal eliminativist, but he does think 'causal' decision theory is better understood as relying on non-casual counterfactuals; likewise Hitchcock (2013) argues that actual causation plays no role in decision theory, though causal dependence, which is closely linked to counterfactuals, does. Finally, van't Hoff (2022) offers a counterfactual decision theory, tailor-made for the causal eliminativist who wants to response to the challenges of how to understand deliberation in the absence of causation. Her decision theory however relies explicitly

4 The Nature of Supplementation

The upshot of the foregoing is that the ground of a distinction between effective and ineffective strategies, one aligning neatly with our commonplace deliberations, cannot be found in basic physics. Assuming we wish to retain something like our present deliberative strategies, we need to supplement that physics in order to recover the distinction. This idea has precedent:

> … we all believe in lots of distinctions physics 'can't see'. Arguably, … all [fundamental physics] needs to describe the events with which it concerns itself are things like tiny particles, gigantic fields, and spacetime. Is there no difference, then, between groups of particles that make up larger wholes, and groups that do not? Should we conclude that, since physics does not mention things like dogs, there is no reason to believe in such things – as opposed to mere swarms of particles arranged in various canine shapes? (Zimmerman 2008, 219)

But there are various approaches to supplementation. These approaches differ on the question of the completeness of physics without causal laws. The *moderate* view takes physics to be complete insofar as it talks about physical entities, but that additional facts about causes and causal laws can be added, consistent with the underlying physics. The *radical* view regards physical and causal laws as making incompatible predictions about physical bodies, and hence physics not only needs causal supplementation, but revision.

Both of these approaches have difficulties. The radical view needs compelling evidence that the predictions of physics are inaccurate, and that where causal and physical predictions disagree, it is the causal predictions that are to be favoured. I know of no such evidence. There are plenty of arguments, of course, that we need to idealize and approximate in order to make use of physical models in local prediction and explanation (Cartwright 1983). But these do not amount to arguments that physical laws have empirical counter evidence when the global state is correctly specified. I'll set aside, then, radical approaches.

Moderate supplementers think that physics is correct so far as it goes. A difficulty for this view comes from the further premise that, when it comes to physical facts, 'as far as it goes' is *everywhere*. That is: in principle, when it comes to prediction and explanation of claims about the motion of material bodies, the existence of a fundamental physical model renders other approaches dispensable, even if they are compatible with the underlying physics. Physics is 'author-

on all three of a partition of the physical states into macrostates, a non-fundamental probability function over possible states, and a similarity metric on possible worlds (van't Hoff 2022, sec. 6): her account of credence in a counterfactual $A > B$ sets it equal to the statistical mechanical probability of B in the current macrostate, in the most similar world in which A holds (which might come apart from the actual statistical mechanical probability of B given A and the current macrostate).

itative when it comes to predicting and describing the production of physical events' (Eagle 2007a, 174). In part this authority is captured by the claim that everything supervenes on the fundamental; fix how things are physically, and you fix how everything is. (That already blocks certain kinds of 'emergentisms' that would press for a robust autonomy for the special sciences.) But it is not just that everything supervenes on the physical: physics is *complete*: every physical event can be wholly described – and when so described, explained and predicted – in purely physical terms. There is, in principle, an exhaustive non-redundant physical account of every event – and as we've argued, that won't be a causal account. It follows, again in principle, that there is no theoretical requirement for causal explanation of any event.

This is obviously only a very rough and sketchy glance at an argument-type that is familiar from discussions of the autonomy of the 'special sciences' that has spawned a very large literature (e.g., Fodor 1974, 1997; Kim 2005; Loewer 2009; Menzies and List 2010; Robertson forthcoming). My argument has a premise about the redundancy of causal information given fundamental physics, so it doesn't immediately generalize to efforts to defend the autonomy of other special science properties and relations.[9] My argument takes its cue from the manifest difficulty of finding a theoretical role for causal regularities in light of the apparent fact that there are no two possible worlds that are wholly alike with respect to fundamental physics and yet unalike in what causes what. A robust defense of causation would involve finding some way to prevent causal relations being trumped by the underlying physics and nevertheless to acknowledge the proper deference of causal connections between particular events to the underlying nomic relations between global states of the world.

5 The Practical Role of Causal Models

The obvious option is to look, not to the metaphysics of causation, but to the epistemic and practical role of causal models. The key issue that then arises is what to say about the view we ought to take of causal models, if they go beyond physics in this non-metaphysical way.

One proposal would emphasise the role of causation in *understanding*. A non-causal physical model of some phenomenon might fail to be explanatory, in the sense of conveying understanding to the explainee, because whether a scientific

9 Barring the premise about the all-encompassing nature of fundamental physics, nothing in my argumnent bears on questions about the autonomy of, e.g., mental properties. There may be excellent reasons to suppose that physics isn't complete when it comes to the description of those physical events that turn out to realize mental states, or thermodynamic states. Admittedly, should the grounds for autonomy of the special sciences involve the autonomy of causal explanations involving special science properties, those arguments may be collaterally damaged by my claim for the in principle dispensibility of causation.

explanation conveys understanding depends only partly on its theoretical content. It depends also in part on the conceptual resources available to the prospective explainee. An explanation that is theoretically redundant might nevertheless have utility if the explainee is not prepared to grasp a more fundamental explanation which trumps it. Doubtless causal concepts are more familiar to us than the equations of fundamental physics, and this proposal is surely correct as far as it goes. But for causation to properly ground a distinction between effective and ineffective strategies it seems like we will need a justification for invoking causal relations that is less tied to idiosyncratic conceptual inabilities of individual knowers.

Loewer suggests that the practical and epistemic rationale for special science models is that they

> characterize aspects of the structure generated by the fundamental physical laws that are especially salient to us and amenable to scientific investigation in languages other than the language of physics. ... the autonomy/irreducibility of the special sciences ... must ultimately be due to facts and laws of micro-physics and to our epistemological situation in the world (Loewer 2009, 222)

Loewer suggests that the challenge of the (metaphysical) completeness of physics should be addressed by examining our standpoint. Moderate supplementers can be understood as offering a variety of attempts to clarify how our epistemic situation supports the use of causal models. They have tried to argue that causal models, even if they are in principle unnecessary, are needed in some more practical or derivative way. Two main families of moderates have presented themselves.

Approximators Some argue that we are justified in making use of causal models because they approximate the real physical explanation under certain conditions. Norton suggests that while causation cannot be recovered from our best science, 'in appropriately restricted circumstances our science entails that nature will conform to one or other form of our causal expectations' (2003, 13). Elga argues that causal models are 'useful to us because so many systems can be treated as isolated from so much of their environments' (Elga 2007, 111). Approximators thus appear to argue that causal models earn their keep through promoting the same goals as physical models – prediction and explanation – and are useful because they can be applied more easily and (under the right circumstances) without appreciable loss of accuracy. But a causal relation that arises only from approximate models is only as real as other entities which are restricted to approximate models: frictionless planes, Newtonian orbits,

Perspectivalists Others argue that causal models serve goals that we have as agents that physical models do not, and that their conditions of adequacy are linked essentially to this agential *perspective*. So, for example, Menzies argues that two physical situations could differ in which causal claims they support while being physical duplicates, due to differences in the context relative to which the causal claims are expressed (2007, 192–93). He ties these differences in context to differences in the course of events that is 'normal or expected or is taken for granted' (Menzies 2007, 220) by the person attributing causation. This context-sensitivity, Menzies thinks, is to be understood as showing that physics needs supplementation by information about such normal courses of events, information that is supplied by our expectations of what will happen if we don't intervene. Menzies tentatively endorses the idea that causal relations are nevertheless real, though in part mind-dependent, given the involvement of expectations. Price (2007) is more full-blooded in endorsing *perspectival realism*, arguing that the deliberative perspective constitutive of our agency, and typical features of what we can know, requires us to make certain assumptions (such as the fixity of the past) that are not required by the physics alone. It is with the support of such assumptions that the outcomes of our interventions are established (Price 2007, secs. 10.7–8). Again, a mind-dependent feature, peculiar to deliberators constituted in roughly the way that we are, is understood to be part of what grounds certain causal claims. Finally, I also endorsed a kind of perspectival realism (Eagle 2007a, 185–87), tied to another aspect of our agential perspective: that we seek understanding and the concomitant ability to generalise our explanations to new scenarios (typically by representing our explanatory conclusions as counterfactual judgments). To do this requires abstracting away from details that might be highly relevant, physically speaking, and adopting a modally rich perspective that involves, once again, certain perspective-dependent assumptions about which details are important, and which are dispensable.[10]

I once found the perspectival approach deeply appealing. It appears to hold out the promise of both respecting the completeness of physics as a theory of things given 'the view from nowhere' (Nagel 1986), while simultaneously supporting the robust reality of causal relations, from the embedded perspective of agents with a particular epistemic position and a circumscribed sphere of influence. It would neatly thread the dilemma we now face. If causal notions aren't fundamentally real, but only real 'from our perspective', we needn't embark on any revision of fundamental physics. But if being real from a perspective is a way of being real,

[10] The approximator/perspectivalist distinction may not be especially hard-and-fast – Norton (2003, 14) says that causes have a 'derivative reality', not so different from the kinds of things perspectivalists say.

we might have a notion of causation that is real enough to vindicate the causal models at the heart of the human sciences. This last contrasts with the approximator's position that the our causal models merely approximate reality, without being in fact accurate; for those who want a more robust independence of the human sciences from any physical basis, the perspectivalist view holds more appeal. But we must be wary of confusing a wish that perspectivalism be true, with an argument that it is.

Causal perspectivalists have some unfinished work to explain exactly what 'real from a perspective' is supposed to come to (Schaffer 2010, 848). One interpretation takes the view to be unashamedly metaphysical, a view perhaps close to Fine's 'relativism', the view that 'Reality is relative to a standpoint; and for different standpoints there will be different realities' (Fine 2005, 262). Such a version of perspectivalism will deny that fundamental physics is authoritative: physics might give us non-relative aspects of reality, but it will omit precisely those facts, like causal facts, that emerge only given an agential perspective. Yet the central conceit of 'relative facts' is hard to accept, especially in this case where (unlike other cases where relativism has been defended, such as tense or the first-person) there is no obvious non-analogical sense of a standpoint being invoked. It would be better to be able to do without radical metaphysics; or perhaps, if one is going to opt for a radical view, a radicalism that is tailored to causation (like that mentioned at the start of this section) seems better motivated.

But this reading of Price's perspectivalism is not perhaps the most plausible; it might be more charitable to read him as committed, like Menzies is, to a kind of contextualism about *cause*. (Indeed, prominent interpretations of Price, like Ismael (2015), read him as committed to an explicit thesis that *cause*, like *local* or *near*, has some covert argument place filled in automatically by a contextually-given perspective.) The contextualist argues that the word *cause* is context-sensitive, so that '*A* causes *B*' could expresses, relative to context *c*, something like '*A* raises the *c*-expected probability of *B*'. The role of context here is to supply just the supplemental ingredient we mentioned at the end of §3, namely, probabalistic or modal information. Contextualism has an advantage over theories which require explicit calculation of the relevant probabilities, if only because context can supply the probabilities to truth conditions directly without awareness on the part of the speaker. But this form of contextualist perspectivalism, while non-mysterious, doesn't address the central worry that the propositions expressed are redundant given the physics: we still need some argument as to why the propositions expressed in context by causal claims can play an interesting role in our conceptual economy given their fundamental dispensibility.

In my earlier paper, I suggested that a *fictionalist* approach might clarify perspectivalism:

> utterances within a given perspective, just as utterances in a fiction, are to be interpreted semantically as if the content of the fiction were true, but not as representing that the content obtains actually. (Eagle 2007a, 188)

On re-reading my earlier discussion, it is far from clear how exactly I thought fictionalism related to perspectivalism, and on the face of it, there is a significant tension between the non-committal ontology of the typical fictionalist and the metaphysical radicalism of the perspectival realist. It is perhaps easier to see how contextualist perspectivalism is compatible with fictionalism. But as contextualism is a theory of truth-conditions – not of whether those conditions obtain – it is difficult to see how contextualist perspectivalism can in fact fulfill the perspectivalist promise to reconcile the completeness of physics with a robust commitment to causation. Perhaps there is a contrast between the context-insensitive language of fundamental physics, and the context-sensitive language of causation – but something more needs to be said about how that contrast could illuminate the metaphysics of causation.

However, while I'm less persuaded than my earlier self of the promise of perspectivalism, I continue to favour fictionalism as an approach to understanding the relation of causation to fundamental physics. In the remainder of this chapter, I want to try and develop the very preliminary remarks I made in 2007 a bit more, and really explain and attempt to justify a fictionalist approach to causation.

Before leaving other moderate supplementing approaches, I want to suggest that fictionalism may be helpful in understanding the approximator's position also. For some approximators may wish to endorse the idea that causal models are models, in the same way that an ideal gas or a frictionless plane is a model: entities that are known to be unreal, but that are used to represent a target system with which they share certain explanatorily potent features. There is a considerable appeal to the idea that scientific models are like fictions, 'creatures of the imagination' as Godfrey-Smith (2009, 101) puts it (see also Contessa 2010; Frigg 2010). I will explore a fictionalist account of causation with some awareness of this potential application, but primarily with the goal of trying to offer an account of our construction and use of causal models that might vindicate our practice in the absence of a reduction to physics (or a Cartwright-style causal inflation of physics).

6 Fictionalism

There are many varieties of fictionalism (Sainsbury 2010; Eklund 2019; Kroon 2011; Caddick Bourne 2013). The fictionalism that interests me is an interpretative approach to an apparently entity-involving discourse D which

> claims that those participating in D should not have (and perhaps, in the case of certain types of discourse, typically *do* not have) truth as their aim when they accept a sentence from D. The norm for acceptance is not truth – one can accept a sentence of D for the best of reasons, justifiably acting on it, using it in one's theorising, drawing inferences and acting on those, and so on, but not actually *believe* it, since there are benefits other than truth for whose sake one should accept such a sentence. (Kroon 2011, 787)

This kind of view gets the name 'fictionalist' because this is an interpretative stance that can often be taken towards fictional discourse. Take the case of fully participatory engagement with fictional texts (Eagle 2007b, 128), e.g., in a seminar on *Bleak House* in which people describe people's actions and speculate on their motivations without hedges like 'according to Dickens' or 'in *Bleak House*'. It is implausible to suppose that these hedges are nevertheless there, covert and unpronounced. Fictionalism about fictional discourse ought not to be pursued using what Lewis termed 'disowning prefixes' (Lewis 2005, 315). That sort of account makes incorrect predictions about how participatory discourse about fictions embeds (R. Joyce 2005, 292–95), and how it interacts with our attitudes (Yablo 2001, 76), and about the flexibility of the conversational role of such utterances (Eagle 2007b, 131–33). So it is preferable to take the content of the sentences at face value.

Instead, fictionalism ought to be pursued by some strategy that allows these sentences to be uttered without being assertions, while looking for all the world as if they are. If there is an explicit 'disowning preface' that signals an entire body of utterances isn't to be taken to have the force of assertion – such as Lewis' example 'I shall say much that I do not believe, starting *now*' – then we needn't look any further for grounds for fictionalism about those utterances. Sometimes the preface only conventionally implicates that the speaker disowns what is to follow; that I take it is the function of 'once upon a time' as a traditional opening to fairy tales.

Most discourses lack any obvious disowning prefix.[11] Fictionalists about a given discourse typically need to figure out *why* someone might disown the entailments of what they say, in the absence of any preface assuring us that, for some reason, they do. Hence fictionalists have tended to attempt to come up with reasons why a given discourse might be worth pursuing in absence of any commitment by speakers to its content. These might be understood as reasons we might have for retaining the discourse once we discover it to be erroneous, or might be understood to be the reasons we've all along implicitly had for adopting that form of discourse in the first place.[12] Either way, identifying such

[11] Lewis (2005, 319) however thinks that even raising anti-realist metaphysical hypotheses about *F*s prior to engaging in apparently *F*-committal discourse can suffice for disowning.

[12] This is close to the distinction between *revolutionary* and *hermeneutic* fictionalism (Burgess and

a purpose for a given discourse is to try to come up with a standard for good and poor contributions to the discourse that is distinct from the truth-normed standards of non-fictionalistically interpreted discourse. Perhaps in the case of participatory criticism in the classroom, the purpose of talking that way is to foster imaginative engagement with the work without the distancing that explicit mention of the fiction would encourage. The relevant norm is not to assert truths, but to make your utterances those that facilitate the achievement of this purpose, by allowing yourself and your classmates to suspend disbelief and bring their considerable skills in folk interpersonal psychology to bear on the fictional situation.

Fictionalism about a domain of discourse isn't committed to thinking that the norm of that discourse is the norm of fictional discourse. There are many possible purposes for a practice, including linguistic practices. Fictionalists needn't think that fictionalism about a domain of discourse has to involve taking its point to be the same as a discourse engaging with a work of fiction (Caddick Bourne 2013, 148). So understood, fictionalists needn't discuss in detail the metaphysics of fictional entities or the structure of operators like 'in *Bleak House*...', except of course insofar as such discussions are helpful in generating suggestions about how to understand their target domain of discourse. Similarly, while the attitude fictionalists recommend taking to a given discourse must fall short of belief, it needn't be restricted to those attitudes, like *pretending*, that have been invoked in the service of understanding fictional discourse.

A further important way in which fictionalists distance themselves from paradigm cases of fictional discourse is in any commitment to an *error theory* about the discourse. There need be no commitment to the falsity of any distinctive claims in the discourse. (In fact, even fictionalists about fiction might wish to reject an error theory for historical fictions, which may contain many true historical details, or certain instances of postmodern 'metatextual' fictions, which may well say true things about themselves.) Some fictionalisms may be motivated by a conviction that some central parts of the discourse cannot be true – moral fictionalists, for example, may be convinced that no properties could function in the way moral properties are expected to. But to say of a discourse that any assertion-like speech acts it involves don't implicate belief, isn't to say that one should believe the contents of those acts to be false.

Given that we have distanced fictionalism from any requirement that distinctively 'literary' phenomena be involved, and we are treating discourses where, semantically, any disavowal of the content is at most present in a tacit preface, fictionalism ends up having a fairly thin characterisation. To be a fictionalist about *F* involves offering an orthodox semantics for *F* discourse that is continuous with the rest of the language, coupled with an unorthodox account of the norms gov-

Rosen 1997; Stanley 2001).

erning assertion-like speech in *F* discourse. Fictionalism is thus a fairly broad church. But without any recourse to the conventions surrounding fiction (like filing a book under 'Fiction' in the library), or to the standard fictional attitudes of pretence or make-believe, would-be fictionalists are confronted with several questions about their approach to *F* discourse.

Disavowal Why should we treat *F* discourse as not requiring truth to be successful, nor as requiring its users to believe in its content?

Retention Why retain that discourse and continue talking of *F*s, rather than stick with truth-normed discourse?

Aim What is the purpose of the discourse in the absence of an aim at truth?

Attitude What attitude should participants in the discourse take or be understood as taking? What is the assertion-like speech act involved?

Answers to these questions will constitute a characterisation of a species of fictionalism about *F*s, and an argument for adopting fictionalism. If we can provide satisfactory answers to the first two questions, that provides a *prima facie* case for a fictionalist approach, at least in the presence of the auxiliary assumption that talk of *F*s should be given an semantics continuous with the rest of natural language. The ultimate case for fictionalism also requires that there be plausible answers to the second two questions. A plausible fictionalism will invoke attitudes that are psychologically realistic, speech acts that are at least arguably attested in other domains, and an aim that is a legitimate goal for a human practice. The questions aren't necessarily to be taken sequentially, however; if we cannot give a reasonable non-cognitivist account of the aim and attitudes involved in a discourse, we might well revisit the question of whether we ought to retain or disavow the discourse in the first place.

Proponents of fictionalism about various subject matters have been more or less dutiful in attempting to answer these questions. So moral fictionalists (Kalderon 2005; R. Joyce 2005) might say: we disavow moral discourse because of its queer metaphysics; we retain moral talk because precommitments to 'exclude from practical deliberation the entertainment of certain options' (R. Joyce 2005, 307) are an effective way to bolster the will to avoid socially and individually detrimental temptations; the purpose of moral discourse is to establish such precommitments; and we take an attitude to a moral claims *M* which is approximately *deferring to M as a guide in action*. Joyce thinks this last attitude is a kind of collective make-believe, where we choose to engage with the fiction of morality in order to establish certain dispositions in ourselves to react in 'the moral way'. Many have questioned whether engaging with an explicit pretence does manage to inculcate the relevant dispositions.

Closer to our concerns is the proto-fictionalism about theoretical entities espoused under the name 'constructive empiricism' by van Fraassen (1980). His answers to our guiding questions are something like this, on my interpretation:

- We needn't endorse (by believing) the content of theoretical entity claims because empiricism ensures that such claims could not amount to knowledge (which norms assertion);
- We ought nevertheless to retain theoretical entity discourse because being 'totally immersed in a scientific world-picture' (van Fraassen 1980, 81) is vital for the successful pursuit of scientific activity, both experimental and theoretical. Science proceeds best when scientists work with theories having claims about unobservable entities as part of their content, rather than working with bowdlerized theories only mentioning observables;
- The aim of science is the construction of 'empirically adequate' theories (theories that are accurate in what they say about observable phenomena); and
- The attitude involved is *acceptance*, amounting to a belief that a theory is empirically adequate,[13] coupled with a practical decision to act more or less as if one believed the theory. The associated speech act seems to be something like assertion within the context of supposing the content of the theory:

> Even if you do not accept a theory, you can engage in discourse in a context in which language use is guided by that theory—but acceptance produces such contexts. (van Fraassen 1980, 12)

A satisfactory fictionalist theory of causation will answer our four questions. Most of these questions can be addressed by taking up points already made in the first half of this chapter. I particularly wish to revisit those earlier arguments in light of van Fraassen's brand of fictionalism. It would not be wildly inaccurate to characterise my proposal here as a kind of constructive empiricism about causation, but where acceptance of causal models is keyed to our deliberative goals rather than scientific inquiry more generally.

We've rehearsed at length the reasons to think the causal talk cannot be given a fully realist treatment: physics is complete §4, but nothing in the physics corresponds to the relations of local determination characteristic of causation §§2, 3. Nothing in these observations tells us that causal models are incompatible with the underlying physics, so we don't have an argument that we have to give up causal representations of the phenomena. But the completeness of physics does suggest that, equally, we cannot be required as a matter of theoretical adequacy to include causal relations in our global models of reality. Representing things

[13] While the attitude of acceptance involves belief in what's said by a sentence with some kind of disavowing prefix – e.g., 'So far as empirical adequacy is concerned, *P*' – the content of what is accepted is just *P*.

causally is optional. In the next section, I'll argue that using causal representations is an option we should exercise; and in the final two sections §§8, 9, that the best way to balance the retetion of causal thought and talk with their optionality is to treat them as a 'useful fiction' (cancelling, of course, the implication that disavowing a causal representation involves regarding it to be false). The point of causal representation isn't to duplicate the function of fundamental physics; and fulfilling the function of fundamental physics isn't all we want from a model of reality.

7 Retaining Causal Talk and Models

Nevertheless, we ought to retain causal talk and our deployment of causal models in situations of human interest. Our earlier discussion also showed that physical models don't give us a good handle on the distinction between effective and ineffective strategies (§2). Interrogate the physics for a potential intervention with respect to some local outcome, and you get back a vast region containing every event of potential physical relevance to that outcome. Changing the character of the whole region is guaranteed to affect the outcome, but is practically infeasible. But the physics gives no guidance about which subregions are the most potentially effective loci of action. But that is precisely the information we want, and that causal models give us.

In Eagle (2007a, 177), I emphasised the importance of causal information for explanation. The way I'd now put the idea is this. The vast network of events of potential physical relevance is altogether too complex and detailed to be explanatory. This is not just because our understanding is too limited to grasp it, but because explanation itself is essentially an abstractive process. To explain an outcome is not to show that its physical antecedents necessitated it. Rather, we explain X when we answer a contrastive question: *why X rather than some salient alternatives X', X'', ...* (van Fraassen 1980, 127). The answer we give must cite some event on which the occurrence of X rather than X' depends: some event such that, had it not happened, some alternative to X would have occurred. Explanatory dependence appears to be a species of counterfactual dependence.

But counterfactual dependence may not suffice, as cases of Simpson's paradox show. Suppose we are considering some statistics about college admissions. The real statistics are a bit more complicated (Bickel, Hammel, and O'Connell 1975), and I'm going to tweak the variables of interest to illustrate my point more clearly. Suppose we collect statistics on admission broken down by whether the applicant applied in the main round to start in the autumn, or for mid-year entry, starting in the spring. We see something like these statistics: overall, the chance of a mid-year entry applicant being admitted is 12.5%, while the chance of a main round applicant being admitted is just 10%. This might suggest that a relevant

explanatory factor of an applicant's chance of admission is time of application: that a main round candidate might have had a higher chance of admission had they applied for mid-year entry instead. This might also suggest certain hypotheses, for example, that admissions committees favour mid-year applicants. But before jumping to policy recommendations, we look at the raw data in a more fine-grained way, broken down by department of application, as follows:

Table 1: Imagined admissions data

	Admit/Apply: English	Admit/Apply: Spanish	Overall
Mid-year	0/15	5/25	5/40
Main round	6/135	10/25	16/160
Overall	6/150	15/50	21/200

What we see is that, though overall main round applicants have a lower admission rate than mid-year applicants, they have a higher admission rate in each department to which they apply. This data suggests many hypotheses about why mid-year applicants apply in this distinctive way, but the hypothesis that they are favoured by admissions committees cannot be sustained. The driver here is that mid-year applicants apply preferentially to more competitive departments: most main round applicants apply to English, with a 4% admission rate, and most mid year applicants apply to Spanish, with a 60% admission rate. This additional information then suggests an opposing counterfactual: that had a main round candidate applied mid-year, they would have had a lower chance of admission.

These counterfactuals have contrary consequents; which is right? Arguably, both – *in the appropriate contexts*. Make the context explicit, and we see that both of these counterfactuals are true:

(1) Had the applicant applied mid-year, they would have had a higher chance of admission – *because had they applied then, they would have more likely applied differentially to less competitive departments*.
(2) Had the applicant applied mid-year, they would have had a lower chance of admission – *because mid-year applicants have lower admission rates, holding fixed the department to which they actually applied*.

Both counterfactuals might be true. But (2) is explanatory in a way (1) is not. This is because it holds fixed a causally relevant factor. As Cartwright notes in her discussion of the real data,

> The difference between the two situations lies in our antecedent causal [assumptions]. We [assume] that applying to a popular department (one with considerably more applicants than positions)

> is just the kind of thing that causes rejection. … If the increased probability for rejection among [main round applicants] disappears when a causal variable is held fixed, the hypothesis of discrimination in admissions is given up… (Cartwright 1979, 433, slightly altered)

By contrast, though (1) it is true and tracks a genuine statistical association, it focuses on a feature – being a mid-year applicant – which we intuitively don't regard as causally significant. We may think that a fixed applicant is likely to have 'intrinsic' preferences between disciplines, but that the question of whether they apply mid-year or not is a more external and haphazard matter. There is no natural kind, 'mid year applicant' – by contrast, 'English applicant' seems a more natural category.

If this is right, then good explanations will invoke counterfactuals that are appropriately backed by causes. It will not do to explain why a given candidate was admitted by citing their status as a mid-year applicant. A better explanation cites the department to which they applied, and recognises that differences between admission rounds in the distribution of applicants to disciplines may play some role, but it is not a straightforwardly causal one. Causation is here essential for discriminating better explanations from worse, even in a broadly counterfactual approach to explanation.

What goes for explanation also goes for deliberation. An agent deciding what to do must balance the expected costs and benefits of various actions. Suppose an applicant is deciding when to apply to graduate school; having just missed the deadline for the main round, they slightly prefer to apply mid-year, rather that wait to the following year. But they vastly prefer being admitted to being rejected. This might summarise their preferences over various outcomes (act/state pairs), assigning numbers – *utilities* – roughly to track something like the degree of relative preferability, as in table 2.

Table 2: Admission decision: matrix of value

Acts	States	
	Admitted	Rejected
Apply mid-year	49	1
Apply next main round	45	0

Evidential Decision Theory says that, given these utilities for outcomes, you ought to choose the action that has the best subjective expected utility. The subjective expected value of an act's utility ('subjective expected utility', for short)

is the credence-weighted average utility of the act across all states. Because the credence of a state isn't independent of the act, we need to use the conditional credence of the state given the act. The general framework is this. Given a potential act A, some background states $S_1, ..., S_n$, someone's utility function U, and their credence function P (a probability function), the expected utility of A is defined:

$$SEU(A) \stackrel{\text{def}}{=} \sum_{i=1}^{n} U(S_i \wedge A)P(S_i \mid A).$$

These numbers lead to decision in accordance with this proposed norm:

EDT You ought to perform the act that has the highest subjective expected utility of those open to you.

What the admissions example shows is that this norm goes awry. Applied uncritically, we might reason as follows:

> We are trying to establish the relative merits of the acts *apply mid-year M* and *apply next year N*, with respect to the outcomes of being admitted (A) or rejected ($\neg A$). The expected utility of applying mid-year is
>
> $$\begin{aligned} SEU(Y) &= U(A \wedge M)P(A \mid M) + U(\neg A \wedge M)(1 - P(A \mid M)) \\ &= 49 \times 5/40 + 1 \times 35/40 = 7. \end{aligned}$$
>
> The expected utility of applying in the next main round is
>
> $$\begin{aligned} SEU(N) &= U(A \wedge N)P(A \mid N) + U(\neg A \wedge N)(1 - P(A \mid N)) \\ &= 45 \times 6/160 = 4.5. \end{aligned}$$
>
> So we ought to apply in the mid-year round.

This looks like bad reasoning in light of the fact that in each department, applicants have better prospects in the main round. The overall statistics are correct, as far as they go, but are misleading, because the increased in probability of mid-year admission is an artefact of the way the statistics are analysed. As Cartwright noted, we ought to think not about whether mid-year application is *correlated* with admission, but with whether it *causally promotes* admission.[14]

Cartwright does not offer a decision theory, exactly. She does offer a definition of 'effective strategy' (1979, 431): S is an effective strategy for G if the ex-

[14] Some versions of EDT (Price 1986, 199; Ahmed 2014) counsel against using these statistics in setting one's credences, and instead advise using the same statistics used by causal decision theory, as below. Whether these versions of EDT are 'really' non-causal, given the reasoning they deploy constraining the choice of appppropriate probability function, and the deflationary version of CDT introduced below, is an interesting question.

pected conditional probability of G given S is higher than the expected unconditional probability of G, where that expectation is calculated over a partition of the outcome space by the causally relevant background factors. Broadly following Skyrms (1980), this idea can be incorporated into a decision theory. This will be a version of *Causal Decision Theory*, though of a sort not so popular these days.

Let $C = \{C_1, ..., C_n\}$ be a *causal partition*: a division of the outcome space into mutually exclusive, jointly exhaustive cells, such that each C_i specifies some way for the relevant causal background factors (those not at the time of decision under the agent's influence) to be distributed. The partition will be one that an agent regards as a good way of thinking about the causal structure, in light of how they represent the situation. There's no requirement that the partition reflect the 'real' causes, whatever those might be. Rather it reflects factors that this agent regards as the background against which they might act. Lewis (1981, 11) suggests that the members of this partition for a given agent could each be a 'maximally specific proposition about how the things [the agent] cares about do and do not depend causally on [their] present actions'. Skyrms (1980, 133) gives a similar characterisation: the cells should be 'maximally specific specifications' of the factors outside the agent's influence (at the time of decision) which are causally relevant to the outcomes. I take it that these descriptions will be satisfied if the cells of the partition are determined by whatever causal model of the decision situation is available to the agent. Given such a partition, let the C-expected credence of O given A, relative to C, be defined:

$$P_C(O \mid A) \stackrel{\text{def}}{=} \sum_{C_i \in C} P(O \mid A \wedge C_i) P(C_i).$$

With this notion in hand, the needed change is simple: agents shouldn't use their credences in outcomes given acts in deliberation: they should use their C-expected credences. The way to implement this is to define a new, causally-sensitive notion of expected utility, to sit alongside SEU (it's just a definition, so can't be right or wrong). The change we need is in our norm; rather than use SEU to evaluate possible acts, we should use this new notion. So let's define the *subjective causally expected utility* by adapting our earlier definition of SEU:[15]

$$CEU(A) \stackrel{\text{def}}{=} \sum_{i=1}^{n} U(S_i \wedge A) P_C(S_i \mid A).$$

In general, of course, $P(O \mid A) \neq P_C(O \mid A)$. Then proposed norm is:

CDT You ought to perform the act that has the highest subjective causally expected utility of those open to you.

[15] This definition is very close to one of Lewis' reformulations of his causal decision theory (1981, 15).

Apply this to our admissions example. The background causal partition in our simple model is given by which department an applicant is to apply to: English (En) or Spanish (Es). Presumably it was once under the agent's control to choose a major; having done so, they are no longer in any position to make a choice about which department to apply to. (This is vital, of course: for it blocks the agent from changing their intended department and thus exploiting the statistical quirk that makes mid-year entry look *prima facie* appealing.)

The relevant probabilities can all be extracted from the frequencies presented in Table 1:

$$P(En) = 150/200 = 3/4; P(Es) = 50/200 = 1/4.$$

Hence

$$\begin{aligned} P_C(A \mid M) &= P(A \mid M \wedge En)P(En) + P(A \mid M \wedge Es)P(Es) \\ &= (0/15 \times 3/4) + (5/25 \times 1/4) = 5/100; \quad \text{and} \\ P_C(A \mid N) &= P(A \mid N \wedge En)P(En) + P(A \mid N \wedge Es)P(Es) \\ &= (6/135 \times 3/4) + (10/25 \times 1/4) = 2/15. \end{aligned}$$

Now we can calculate CEU for our possible acts:

$$\begin{aligned} CEU(M) &= U(A \wedge M)P_C(A \mid M) + U(\neg A \wedge M)(1 - P_C(A \mid M)) \\ &= 49 \times 5/100 + 1 \times 95/100 = 3.4; \quad \text{and} \\ CEU(N) &= U(A \wedge N)P_C(A \mid N) + U(\neg A \wedge N)(1 - P_C(A \mid N)) \\ &= 45 \times 2/15 = 6. \end{aligned}$$

The final step is to apply our new causal norm, which (in disagreement with EDT) recommends that you ought to apply in the main round, regardless of your chosen subject. The credences involved in rational decision should be causally informed, not merely reflecting the statistics, but reflecting how you as an agent are positioned with respect to those statistics. The bare fact that mid-year entrants have a higher success rate *simpliciter* doesn't reflect the decision facing the agent. They are trying to decide what to do, holding fixed what they are assuming is no longer under their control, if it ever was – such as to which program they might be applying.

Most recently popular versions of Causal Decision Theory follow Gibbard and Harper (1978) in calculating expected utility using credences in causally-interpreted counterfactuals (Lewis 1981, 21–28; J. M. Joyce 1999; Hitchcock 2013), rather than using causally-informed credences. The counterfactual approach is however beset by various controversies and challenges, about the logic of counterfactuals and their interaction with conditional probabilities, that the causally-informed credence approach neatly sidesteps. The causal partition is of course rich in information that entails counterfactual claims (or at least, constrains the propositions they express relative to various contexts), and it may

well be possible to extract the Gibbard and Harper version of Causal Decision Theory from the Skyrms/Lewis/Cartwright theory. The version of CDT sketched here has the virtue of making the role of causal information and causal models especially prominent.

This has been a long excursion through the details of causal explanation and causal decision making. The details matter, because they show that appeal to causal information is unavoidable if we want theories of explanation and rational decision that gives correct verdicts about what to think or do in various clear cases, and hence might be a viable candidate to extend as a general purpose approach to explanation and deliberation. The overall argument for retention of causal notions is that our notions of rational choice and explanation are thoroughly permeated by irreducible causal notions. Causal information, drawn from our pre-theoretical background assumptions or (preferably) from properly tested causal models, is a precondition for explanation and deliberation to proceed in anything like the way we think they do. Conversely, if there were no agents in need of understanding, or no one needing to make choices, there may be no need to invoke causal models for any other purpose. Causal information is an essential precondition to taking a deliberative stance, but nothing in the physical laws necessitates that any agents take that stance.

The role of causal information in choice and explanation lies somewhat in the background. In the version of Causal Decision Theory sketched above, it plays a role in the delineation of a causal partition. In the counterfactual theory of explanation, it plays a role in characterising which contexts are such that true counterfactuals in them are explanatory – namely, those contexts in which the information held fixed under counterfactual assumptions is causally relevant to what needs explanation.

In neither case is the truth of the causal information at issue in the deliberative or explanatory project.[16] In deliberation, what is centrally at issue for an agent is what to do; their decision turns on the actions available to them, and the impact of those potential actions on the outcomes they value. The framework above involves the agent antecedently making causal assumptions about how their actions lead to outcomes. But what is at issue in a claim about what it is rational for someone to do – what two rational agents might be centrally disagreeing about if they reach different verdicts about the right course of action – is the values they espouse and the credences they have. The broadly presuppositional role of causal information in explanation and deliberation means that the truth or falsity of causal claims is largely beside the point for agents. Having

[16] In semantics, 'at issue' content is what is centrally said by a sentence, rather than incidental to, or presupposed or implied by it (Potts 2015).

causal presuppositions is enough to get the deliberative project off the ground; those presuppositions provide a framework for having an agential perspective.

Eliminativists about causation, like Russell, will deny the truth of these presuppositions, and argue that the deliberative project ought to proceed in some radically different way, relying only on physical structure, and not on any proposed 'excess' causal structure. I see no real prospect that this reconstructive ambition will be satisfied. Consider an attempt to pursue it within the decision theory sketched above. The background partition cannot be a causal one any longer, but must rather include background states of physical relevance that are not under the agents control. There are only two ways to proceed here.

- We can include every physically relevant background condition. This will give us an extremely fine-grained partition; indeed, under determinism, the partition will assign each possible physical model its own cell. This will enable us to define a notion of physically expected credence – *if* agents have credences over this extremely rich partition. Even if that unrealistic assumption is met, the partitions are so fine-grained that the background partition, for any proposed action A, already entails A. The physically expected credence of an outcome given an action reduces to just the expected credence of that outcome. When this credence is plugged into the CEU formula, we get the triviality that the expected utility of an action is just the expected value of a physical state – action falls out of the picture entirely, and this is no longer a theory of deliberation (rather it is a theory about what physical process you should hope take a course through your body and environment). As Price (2007, 281) puts it: a deliberator must think of 'her own actions as ... not themselves determined by anything "further back" '. The fine-grained physical partition is such that each cell entails an action (or its negation). Accordingly the agent who uses this partition isn't really deliberating, on Price's view: they have credences in which bodily movements they'll come to perform, and they have hopes and wishes about those prospects, but they can't be deciding, as that involves the agent making it such that a certain action occurs, as it were 'independently' of its physical precursors. The deliberator whose background partition is fine-grained enough won't be able to ignore the fact that this picture of choice simply doesn't seem to be reconciled with the existence of a complete physical account of each bodily movement.[17]
- If we don't include every physically relevant background condition, we are back to the original admissions case with which we began, where spurious

[17] Van Fraassen makes the related point that an omniscient being is not in the business of seeking explanations; there is necessarily some ignorance involved in any explanation – as well as some background assumptions taken for granted which determines the kind of explanation sought (van Fraassen 1980, 130).

> associations between variables which are physically real must be regarded as candidates to guide credence in decision, as in EDT. The existence of intuitively 'better' choices of background variables is irrelevant in the absence of physical grounds to privilege those variables. If we somehow try to make use of all possible selections of potential background information, our decision theory will be worryingly non-committal about which actions are rationally preferable.

Neither of these approaches is satisfactory for us. Either we aren't offering an account that helps us make decisions, or we offer an account which is subject to counterexample. The invocation of causal information makes for a useful and plausible decision theory.

The case for retention of causal models in the face of their redundancy in the face of physics boils down to the fact that for agents like us, causal models provide non-redundant background information that is vital for intuitively defensible choices and explanations. The content of the 'causal fiction' is something like this: that there is a privileged way of partitioning the space of outcomes that reflects the background causal structure. Adopting this fiction overtly, or implicitly presupposing it, is a necessary precondition for us to do anything that resembles deliberation and explanation.

This case for retention of causal notions is basically that we couldn't rationally deliberate without them. But we also need some guarantee that rational deliberation is a viable strategy; that our world isn't fundamentally hostile to limited agents like those we take ourselves to be. The case that the world is *approximately* causal would provide such a guarantee, vindicating a deliberative stance as a good trade-off between accuracy and implementation. We discussed part of this case in §3 when discussing the fact that, by and large, many physical systems can be treated as if they were (quasi-)isolated, and hence approximated by a local causal model.

8 The Aim of Causal Models

According to van Fraassen, there is no prospect of empirical knowledge of theoretical entities. Norms governing assertion then forbid us from asserting claims about theoretical entities. What then is the rationale for continuing to make, in an assertion-like way, claims about theoretical entities? Van Fraassen suggests: full immersion in theoretical entity discourse facilitates achieving the aims of the science, namely, the development of empirically adequate theories. Van Fraassen is, on my reading, a causal fictionalist, as causal relations are unobservable, posited to explain observed correlations and patterns. We could follow van Fraassen and argue that the aim of a causal model is just the aim of science more generally: empirical adequacy.

This wouldn't be a wholly satisfactory answer for us. For one thing, the arguments above don't involve anti-realism about theoretical science; causal fictionalism is compatible with robust realism about physics. For another, as just discussed at length, causal models seem to play a distinctive role in our cognitive economy. They are retained not because they help us with 'science' in general, but with particular applications of scientific knowledge. A good account of the aims of causal models, and the human sciences more generally, would explain how invoking causal notions in deliberation and explanation conduces to that aim.

The obvious thought given what we've already argued is that the aim of constructing causal models is to enable effective deliberation and explanation. I've already suggested that without causal assumptions, we may not even get activities that are recognisable to us as deliberative. So in order to facilitate our competent participation in an activity that is fundamental to limited agents like ourselves, we need to make assumptions about the background fixed causal structure against which we consider our options and evaluate candidate explanations.

Our causal models need not be true to play this role. But even very modest externalistic constraints on explanation and deliberation must say that rational choice and successful explanation cannot float wholly free from worldly matters. (Perhaps some literary fictions are measured by wholly internal standards of achievement, with their aims not involving truth even in part, but I don't think the causal fiction could be like that.) Just as the constructive empiricist says that successful science should be true in what it entails about observable matters, the causal fictionalist should I think say that successful causal models ought to be *physically adequate*: true in what they entail about physical matters. The causal relations they involve is excess structure from the perspective of physics, but the correlations and patterns of occurrence between variables that a causal model entails should ideally agree with what an ideal physics would entail about the situation.

This means that Humean physical patterns, such as relative frequencies and regularities about associations between event types, must be accurately predicted by a causal model that is successfully meeting the aims of causal modelling. This means at least that:

1. The pattern of probabilistic associations in a causal model must be statistically accurate to the observed frequencies (§1). This notion of accuracy doesn't require perfect match with observed frequencies, only that the theoretical probabilities should fit the frequency evidence.
2. Causal relations support counterfactuals; these must agree with the results any experimental interventions designed to test dependencies (such as RCTs), as well as agree with any counterfactual dependencies that follow from the underlying physics under certain idealizing approximative

assumptions. For example, the assumption of physical isolation or quasi-isolation allows us to extract counterfactuals from physical models, because it in effect characterises some possible differences between models as gratuitous, and facilitates our focus on a subclass of models that then fix what would happen under a certain intervention (§3). So we certainly want any acceptable causal model to be true in what it says about isolated local systems.

Those are empirical constraints on causal models, helping us decide which models we might accept. We also want to evaluate their outcomes, in the sense that a good causal model should facilitate effective choice, at least in the long run, by an agent's own lights. So any successful causal model M should be such that an agent whose credences are well-regulated (either well-calibrated (van Fraassen 1983) or well-matched to their expectation of the chances (Lewis 1986)), and who accepts a causal partition based on M, should in the long run not do systematically worse than they would have if they had relied on some other candidate causal model. That is, adequate causal models must be among the best available to the agent in terms of serving their deliberative ends. This conception of adequacy means that an adequate causal model can come to be inadequate if better rivals are constructed, so perhaps we might say that a model is inadequate if there is some extant rival such that in the long run those who use the rival to set their credences tend to achieve valuable outcomes (according to them).

9 The Fictionalist Attitude

I have already indicated that fictionalists need not take an attitude of 'making believe' or pretence to the content of the fiction (§6). Given what I've said, especially about the role causal models play in practice, we need an attitude that is not wholly non-committal, but which is such that taking that attitude to different models could lead to different actions.

Van Fraassen suggests an attitude he calls *acceptance* as the appropriate one to take to scientific theories:

> to accept a theory is (for us) to believe that it is empirically adequate—that what the theory says *about what is observable* (by us) is true. (van Fraassen 1980, 18)

This is the cognitive component of acceptance; to accept P just is to believe some weaker proposition $\mathcal{A}p$. But to accept a theory is linked with practice in a way that belief in the weaker proposition is not, because it is also associated with a practical commitment to deploy the theory in scientific reasoning (rather than to deploy only the claim that things are empirically just as if it were true).

It involves 'a commitment to confront any future phenomena by means of the conceptual resources of this theory' (van Fraassen 1980, 12).

Of course causal fictionalists would be well-advised not to use van Fraassen's notion, because many causal models with merely unobservable differences might nevertheless lead to different counterfactual predictions, and hence different recommendations in action. But there is an obvious candidate attitude in the vicinity, given what we've said about the aims of causal science:

Guidance To be *guided* by a theory is to believe that it physically adequate, and to undertake to frame decision problems and explanations using causal and counterfactual information provided by the theory.

This commitment is implemented in a distinctive way in our discussion above. Recall that causal propositions are not directly the objects of credence in our framework. Rather, causal propositions establish the background structure, either selecting a privileged partition of the outcome space or structuring what is held fixed under counterfactual suppositions. In neither case are we invited to consider any propositions with 'causation' as a constituent. In the admissions example, the partition was over a family of propositions about the department applied to – not over propositions about how the department applied to is causally related to your admissions prospects. Likewise in explanation; the explanatory counterfactual is the one evaluated in a context in which the department applied to is held fixed because it is a causally relevant factor – but no propositions about causally relevant factors are themselves objects of credence. The notion of guidance above fits this neatly. That consequences of coming to be guided by a theory aren't exhausted by changes in one's cognitive attitudes, but also involve the choices about framing one makes, which may not appear in one's credences in any obvious way. Full belief too might lead one to the same sort of framing, but that would differ from guidance precisely on the question of belief in the theory.

To engage quasi-assertorically in causal discourse is to express your commitments. To say, 'which department you are applying to is among the contributing causes of whether you will be admitted' is, on this view, to express your commitment to a certain causal partition, and (also) a recommendation to your hearers to adopt the same commitment. According to Joyce, moral fictionalists make moral claims to bolster, at an individual and group level, the precommitments those moral claims embody. Likewise, causal fictionalists make causal claims to endorse certain structural assumptions as approvable in practical activities.

This may be contentious, as this kind of expressive function for causal language may seem to be invoking a novel speech act not elsewhere attested. However, I wonder if this is actually a by-product of too narrow a conception of assertion. If the causal fictionalist is right, there are two different kinds of attitudes

that both broadly endorse the content of a claim – believing/having high credence in the claim, and adopting that claim to scaffold one's beliefs/credences. If assertion is (roughly) the speech act one opts for in wanting to endorse the content of what one says, then it may be that some assertions all along were attempts to endorse their contents as apt choices to scaffold one's credal structure. This idea would require further development, but I'm sufficiently assured by it to be confident that there is a defensible interpretation of causal discourse as not committing all of its participants to overt belief in causal claims.

10 Conclusion

Fictionalist proposals are very alluring to metaphysicians. They hold out the promise of avoiding commitment to problematic entities while keeping the benefits of talking about them. After an initial flurry of interest in fictionalism about possible worlds (Rosen 1990), numbers (Yablo 2001), and morality, the discussion waned somewhat. No doubt this is partly to do with philosophical sociology and fashion – perhaps the rise of grounding? (Which has itself recently been offered a fictionalist treatment: Thompson (2021).) But it is also due to the difficulty of spelling out decent grounds for scepticism about *F*s while making a powerful case that *F* talk yields irreplaceable benefits. I think the discussion above illustrates this; the long subsection on retention (§7) was largely devoted to trying to show that non-causal explanation and decision was insufficient on its own. That section was hard but I think necessary work before we can reap the purported benefits of fictionalism.

One major issue tempts further inquiry. We've seen that the indispensability of causation is intimately bound up with counterfactual and probabilistic information. If we've been lead to fictionalism about causation, ought we, for consistency's sake, also be fictionalists about modality? Some time ago, Stalnaker said:

> Sometimes I am tempted to believe that there is only an actual world. But we do represent to ourselves pictures of ways that things might be, or might have been, and this practice is not just the idle exercise of our imaginations; it is central to some of our more serious activities such as giving scientific explanations of how and why the *actual* world works the way it does.(Stalnaker 1979, 354).

This might form the germ of a modal fictionalism that really deserves the name: a view on which modal distinctions themselves are artefacts of a representational stance that is pragmatically unavoidable for us. Such a project is radical enough to make causal fictionalism seem mundane, but would seem to be motivated by some of the same concerns leading us to causal fictionalism here.[18]

[18] Thanks to an anonymyous referee and Yafeng Shan for comments, and particularly to Yafeng

References

Ahmed, Arif. 2014. *Evidence, Decision and Causality*. Cambridge University Press.

Albert, David Z. 2013. "Wave Function Realism." In *The Wave Function: Essays on the Metaphysics of Quantum Mechanics*, edited by Alyssa Ney and David Z Albert, 52–57. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199790807.003.0001.

Bickel, P. J., E. A. Hammel, and J. W. O'Connell. 1975. "Sex Bias in Graduate Admissions: Data from Berkeley." *Science* 187 (4175): 398–404. https://doi.org/10.1126/science.187.4175.398.

Burgess, John P, and Gideon Rosen. 1997. *A Subject with No Object*. Oxford: Oxford University Press.

Caddick Bourne, Emily. 2013. "Fictionalism." *Analysis* 73 (1): 147–62. https://doi.org/10.1093/analys/ans126.

Cartwright, Nancy. 1979. "Causal Laws and Effective Strategies." *Noûs* 13: 419–37. https://doi.org/10.2307/2215337.

———. 1983. *How the Laws of Physics Lie*. Clarendon Press.

———. 1989. *Nature's Capacities and Their Measurement*. Oxford University Press.

———. 1999. *The Dappled World: A Study of the Boundaries of Science*. New York, NY: Cambridge University Press.

Cartwright, Nancy, and Jeremy Hardie. 2012. *Evidence-Based Policy: A Practical Guide to Doing It Better*. Oxford: Oxford University Press.

Contessa, Gabriele. 2010. "Scientific Models and Fictional Objects." *Synthese* 172 (2): 215–29. https://doi.org/10.1007/s11229-009-9503-2.

Curiel, Erik. 2021. "Singularities and Black Holes." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N Zalta, Fall 2021. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/fall2021/entries/spacetime-singularities/.

Dawes, Gregory W. 2017. "Ancient and Medieval Empiricism." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N Zalta, Winter 2017. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/win2017/entries/empiricism-ancient-medieval/.

Demarest, Heather, and Michael Townsen Hicks. 2021. "Isolation, Not Locality." *Philosophy and Phenomenological Research* 103 (3): 607–19. https://doi.org/10.1111/phpr.12731.

Dorr, Cian. 2016. "Against Counterfactual Miracles." *Philosophical Review* 125 (2): 241–86. https://doi.org/10.1215/00318108-3453187.

Eagle, Antony. 2007a. "Pragmatic Causation." In *Causation, Physics and the*

for his forebearance. I wish also to acknowledge the support of the Australian Research Council under grant DP200100190.

Constitution of Reality: Russell's Republic Revisited, edited by Huw Price and Richard Corry, 156–90. Oxford University Press.

———. 2007b. "Telling Tales." *Proceedings of the Aristotelian Society* 107 (2): 125–47. https://doi.org/10.1111/j.1467-9264.2007.00215.x.

Earman, John. 1986. *A Primer on Determinism*. Vol. 32. D. Reidel.

Eklund, Matti. 2019. "Fictionalism." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N Zalta, Winter 2019. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/win2019/entries/fictionalism/.

Elga, Adam. 2007. "Isolation and Folk Physics." In *Causation, Physics and the Constitution of Reality: Russell's Republic Revisited*, edited by Huw Price and Richard Corry, 106–19. Oxford University Press.

Field, Hartry. 2003. "Causation in a Physical World." In *Oxford Handbook of Metaphysics*, edited by Michael J Loux and Dean W Zimmerman, 435–60. Oxford: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199284221.003.0015.

Fine, Kit. 2005. "Tense and Reality." In *Modality and Tense*, 261–320. Oxford University Press.

Fisher, R A. 1935. *The Design of Experiments*. Oliver; Boyd.

Fodor, Jerry. 1974. "Special Sciences, or the Disunity of Science as a Working Hypothesis." *Synthese* 28 (2): 97–115. https://doi.org/10.1007/bf00485230.

———. 1997. "Special Sciences: Still Autonomous After All These Years." *Philosophical Perspectives* 11 (June): 149–63. https://doi.org/10.1111/0029-4624.31.s11.7.

Frigg, Roman. 2010. "Models and Fiction." *Synthese* 172 (2): 251–68. https://doi.org/10.1007/s11229-009-9505-0.

Frisch, Mathias. 2014. *Causal Reasoning in Physics*. Cambridge University Press.

Gibbard, Allan, and William L Harper. 1978. "Counterfactuals and Two Kinds of Expected Utility." In *Foundations and Applications of Decision Theory*, edited by C A Hooker, J J Leach, and E F McClennan, 1:125–62. Dordrecht: D. Reidel.

Godfrey-Smith, Peter. 2009. "Models and Fictions in Science." *Philosophical Studies* 143 (1): 101–16. https://doi.org/10.1007/s11098-008-9313-2.

Granger, C. W. J. 1969. "Investigating Causal Relations by Econometric Models and Cross-Spectral Methods." *Econometrica* 37 (3): 424–38. https://doi.org/10.2307/1912791.

Hesslow, Germund. 1976. "Two Notes on the Probabilistic Approach to Causality." *Philosophy of Science* 43: 290–92. http://www.jstor.org/stable/187270.

Hill, A Bradford. 1965. "The Environment and Disease: Association or Causation?" *Proceedings of the Royal Society of Medicine* 58 (May): 295–300. https://doi.org/10.1177/003591576505800503.

Hitchcock, Christopher. 2001. "The Intransitivity of Causation Revealed in Equa-

tions and Graphs." *The Journal of Philosophy* 98 (6): 273–99. http://www.jstor.org/stable/2678432.

———. 2007. "What Russell Got Right." In *Causation, Physics and the Constitution of Reality: Russell's Republic Revisited*, edited by Huw Price and Richard Corry, 45–65. Oxford University Press.

———. 2013. "What Is the 'Cause' in Causal Decision Theory?" *Erkenntnis* 78 (S1): 129–46. https://doi.org/10.1007/s10670-013-9440-9.

Hoover, Kevin D. 2008. "Causality in Economics and Econometrics." In *The New Palgrave Dictionary of Economics*, edited by Steven N. Durlauf and Lawrence E. Blume, 1–13. Palgrave Macmillan UK. https://doi.org/10.1057/978-1-349-95121-5_2227-1.

Howick, Jeremy, Iain Chalmers, Paul Glasziou, Trish Greenhalgh, Carl Heneghan, Alessandro Liberati, Ivan Moschetti, et al. 2011. "The Oxford 2011 Levels of Evidence." Oxford Centre for Evidence-Based Medicine. http://www.cebm.net/index.aspx?o=5653.

Ismael, Jenann. 2015. "How Do Causes Depend on Us? The Many Faces of Perspectivalism." *Synthese* 193 (1): 245–67. https://doi.org/10.1007/s11229-015-0757-6.

Joyce, James M. 1999. *The Foundations of Causal Decision Theory*. Cambridge University Press.

Joyce, Richard. 2005. "Moral Fictionalism." In *Fictionalism in Metaphysics*, edited by Mark Eli Kalderon, 287–313. Oxford: Oxford University Press.

Kalderon, Mark Eli. 2005. *Moral Fictionalism*. Oxford University Press.

Kim, Seahwa. 2005. "Modal Fictionalism and Analysis." In, edited by Mark Eli Kalderon, 116–33.

Kroon, Frederick. 2011. "Fictionalism in Metaphysics." *Philosophy Compass* 6 (11): 786–803. https://doi.org/10.1111/j.1747-9991.2011.00442.x.

Latham, Noa. 1987. "Singular Causal Statements and Strict Deterministic Laws." *Pacific Philosophical Quarterly* 68 (1): 29–43.

Levitz, Lauren, Mark Janko, Kashamuka Mwandagalirwa, Kyaw L. Thwai, Joris L. Likwela, Antoinette K. Tshefu, Michael Emch, and Steven R. Meshnick. 2018. "Effect of Individual and Community-Level Bed Net Usage on Malaria Prevalence Among Under-Fives in the Democratic Republic of Congo." *Malaria Journal* 17 (1). https://doi.org/10.1186/s12936-018-2183-y.

Lewis, David. 1973. *Counterfactuals*. Oxford: Blackwell.

———. 1981. "Causal Decision Theory." *Australasian Journal of Philosophy* 59 (1): 5–30. https://doi.org/10.1080/00048408112340011.

———. 1986. "A Subjectivist's Guide to Objective Chance." In *Philosophical Papers*, 2:83–132. Oxford: Oxford University Press.

———. 2005. "Quasi-Realism Is Fictionalism." In *Fictionalism in Metaphysics*, edited by Mark Eli Kalderon, 314–21. Oxford: Oxford University Press.

Lindstrom, Lamont. 1993. *Cargo Cult*. University of Hawai'i Press. https://doi.org/10.2307/j.ctv9zcktq.

Loewer, Barry. 2009. "Why Is There Anything Except Physics?" *Synthese* 170 (2): 217–33. https://doi.org/10.1007/s11229-009-9580-2.

Lucas, Robyn M, and Anthony J McMichael. 2005. "Association or Causation: Evaluating Links Between 'Environment and Disease'." *Bulletin of the World Health Organization* 83 (10): 792–95.

Maudlin, Tim. 2021. "Relativity and Space-Time Geometry." In *The Routledge Companion to Philosophy of Physics*, edited by Eleanor Knox and Alastair Wilson. Routledge.

Menzies, Peter. 2007. "Causation in Context." In *Causation, Physics and the Constitution of Reality: Russell's Republic Revisited*, edited by Huw Price and Richard Corry, 191–223. Oxford University Press.

Menzies, Peter, and Christian List. 2010. "The Causal Autonomy of the Special Sciences." In *Emergence in Mind*, edited by Cynthia McDonald and Graham McDonald, 108–29. Oxford: Oxford University Press.

Mill, John Stuart. (1874) 1974. *A System of Logic, Ratiocinative and Inductive, Books I–III*. Edited by John M Robson. 8th ed. The Collected Works of John Stuart Mill. University of Toronto Press; Routledge & Kegan Paul.

Nagel, Thomas. 1986. *The View from Nowhere*. Oxford: Oxford University Press.

Ney, Alyssa. 2015. "Fundamental Physical Ontologies and the Constraint of Empirical Coherence: A Defense of Wave Function Realism." *Synthese* 192 (10): 3105–24. https://doi.org/10.1007/s11229-014-0633-9.

Norton, John D. 2003. "Causation as Folk Science." *Philosophers' Imprint* 3 (4): 1–22. http://hdl.handle.net/2027/spo.3521354.0003.004.

Paul, L A, and Ned Hall. 2013. *Causation: A User's Guide*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199673445.001.0001.

Pearl, Judea. 2000. *Causality: Models, Reasoning and Inference*. Cambridge University Press.

Peterson, Martin. 2017. *An Introduction to Decision Theory*. 2nd ed. Cambridge University Press. https://doi.org/10.1017/9781316585061.

Potts, Christopher. 2015. "Presupposition and Implicature." In *The Handbook of Contemporary Semantic Theory*, edited by Shalom Lappin and Chris Fox, 2nd ed., 168–202. Wiley-Blackwell. https://doi.org/10.1002/9781118882139.ch6.

Price, Huw. 1986. "Against Causal Decision Theory." *Synthese* 67 (2): 195–212. https://doi.org/10.1007/bf00540068.

———. 2007. "Causal Perspectivalism." In *Causation, Physics and the Constitution of Reality: Russell's Republic Revisited*, edited by Huw Price and Richard Corry, 250–92. Oxford University Press.

Ramsey, F P. (1929) 1990. "General Propositions and Causality." In *Philosophical Papers*, edited by D H Mellor, 145–63. Cambridge: Cambridge University Press.

Robertson, Katie. forthcoming. "Autonomy Generalised; or, Why Doesn't Physics Matter More?" *Ergo*, forthcoming.

Rosen, Gideon. 1990. "Modal Fictionalism" 99: 327–54. http://www.jstor.org/stable/2255102.

Russell, Bertrand. 1913. "On the Notion of Cause." *Proceedings of the Aristotelian Society* 13 (1): 1–26. https://doi.org/10.1093/aristotelian/13.1.1.

Sainsbury, R M. 2010. *Fiction and Fictionalism*. Routledge.

Schaffer, Jonathan. 2010. "Causation, Physics, and the Constitution of Reality: Russell's Republic Revisited, Edited by Huw Price and Richard Corry." *Mind* 119 (475): 844–48. https://doi.org/10.1093/mind/fzq068.

Simon, Herbert A. 1977. "Causal Ordering and Identifiability." In *Models of Discovery*, 53–80. Boston Studies in the Philosophy of Science, LIV. Springer Netherlands. https://doi.org/10.1007/978-94-010-9521-1_5.

Skyrms, Brian. 1980. *Causal Necessity*. Yale University Press.

Solomonoff, Ray. 1964. "A Formal Theory of Inductive Inference, Part I." *Information and Control* 7: 1–22. https://doi.org/10.1016/S0019-9958(64)90223-2.

Spirtes, Peter, Clark Glymour, and Richard Scheines. 2000. *Causation, Prediction and Search*. MIT Press.

Stalnaker, Robert. 1979. "Anti-essentialism." *Midwest Studies In Philosophy* 4 (1): 343–55. https://doi.org/10.1111/j.1475-4975.1979.tb00385.x.

Stanley, Jason. 2001. "Hermeneutic Fictionalism." *Midwest Studies In Philosophy* 25: 36–71. http://www.blackwell-synergy.com/doi/abs/10.1111/1475-4975.00039.

Stovitz, Steven D, and Ian Shrier. 2019. "Causal Inference for Clinicians." *BMJ Evidence-Based Medicine* 24 (3): 109–12. https://doi.org/10.1136/bmjebm-2018-111069.

Suppes, Patrick. 1970. *A Probabilistic Theory of Causality*. North-Holland.

Thompson, Naomi. 2021. "Setting the Story Straight: Fictionalism about Grounding." *Philosophical Studies*, June. https://doi.org/10.1007/s11098-021-01661-w.

van Fraassen, Bas C. 1980. *The Scientific Image*. Oxford: Clarendon Press. https://doi.org/10.1093/0198244274.001.0001.

———. 1983. "Calibration: A Frequency Justification for Personal Probability." In *Physics, Philosophy and Psychoanalysis*, edited by Robert S Cohen and Larry Laudan, 76:295–319. Boston Studies in the Philosophy of Science. D. Reidel.

van't Hoff, Alice. 2022. "In Defense of Causal Eliminativism." *Synthese* 200 (393). https://doi.org/10.1007/s11229-022-03875-9.

Wallace, David, and C G Timpson. 2010. "Quantum Mechanics on Spacetime I: Spacetime State Realism." *The British Journal for the Philosophy of Science* 61 (4): 697–727. https://doi.org/10.1093/bjps/axq010.

Williamson, Jon. 2019. "Establishing Causal Claims in Medicine." *International Studies in the Philosophy of Science* 32 (1): 33–61. https://doi.org/10.1080/0269

8595.2019.1630927.
Woodward, James. 2003. *Making Things Happen*. Oxford University Press.
Woodward, Jim. 2002. "What Is a Mechanism? A Counterfactual Account." *Philosophy of Science* 69 (S3): S366–77. https://doi.org/10.1086/341859.
Yablo, Stephen. 2001. "Go Figure: A Path Through Fictionalism." *Midwest Studies In Philosophy* 25 (1): 72–102. https://doi.org/10.1111/1475-4975.00040.
Zimmerman, Dean W. 2008. "The Privileged Present: Defending an 'A-theory' of Time." In *Contemporary Debates in Metaphysics*, edited by Theodore Sider, John Hawthorne, and Dean W Zimmerman, 211–25. Blackwell.