Autonomy generalised; or, Why doesn't physics matter more?

Katie Robertson

forthcoming in Ergo

November 18, 2021

Abstract

In what sense are the special sciences autonomous of fundamental physics? Autonomy is an enduring theme in discussions of the relationship between the special sciences and fundamental physics or, more generally, between higher and lower-level facts. Discussion of 'autonomy' often fails to recognise that autonomy admits of degrees; consequently, autonomy is either taken to require full independence, or risk relegation to mere apparent autonomy. In addition, the definition of autonomy used by Fodor, the most famous proponent of the autonomy of the special sciences, has been robustly criticised by Loewer.

In this paper I develop a new account of autonomy following Woodward (2018) which I dub 'generalised autonomy' since it unifies dynamical, causal and nomic autonomy. Autonomy, on this account, can be partial: some lower-level details matter while others do not. To summarise: whilst the detailed lower level is unconditionally relevant, conditionalising on the higher-level facts renders some lower-level details irrelevant. The macrodependencies that the higher-level facts enter into — be they dynamical, causal or nomic — screen off the underlying microdetails.

This account helps resolve an explanatory puzzle: if the lower-level facts in some way underpin the higher-level facts, why don't the lower-level details matter more for the day-to-day practice of the special sciences? The answer will be: the facts uncovered by the special sciences are autonomous in my sense, and so practitioners of these special sciences need not study more fundamental sciences, since these underlying facts are genuinely (albeit conditionally) irrelevant.

1 Introduction

Why care about how we define 'autonomy'? Autonomy is invoked to emphasise the importance and independence of the special sciences. The motivations for emphasising this importance and independence range from the metaphysical to the mercenary.

Fodor (1974) famously argues that the special sciences are autonomous, and that this supports his metaphysical position: non-reductive physicalism. Yet Loewer (2009) has argued that the success of Fodor's project depends sensitively on how autonomy is characterised. In contrast, Anderson's discussion of autonomy in his seminal 'More is different' paper was motivated, in part, by more practical interests (Anderson, 1972). In terms of funding, solid state physics trailed far behind its more glamorous cousin, particle physics, with its expensive accelerators. By emphasising how the special sciences, such as condensed matter physics, are both importantly different from—and independent of—fundamental, i.e. high-energy, physics Anderson made a case for redressing the funding and prestige gap faced by solid state physics (Martin, 2015).

What are my motivations for defining an account of autonomy? Sometimes we have a range of scientific descriptions for a given system.[1] Sometimes this is an 'in principle' matter: for example, the Schrödinger equation in principle describes the wildebeest migrating across the Serengeti, although few dare to hope that such a description could be written down. But sometimes we have these different descriptions in hand: we can describe a gas in a box according to thermodynamics, statistical mechanics, quantum mechanics or classical mechanics. The thermodynamical facts about the gas are higher-level facts than the lower-level facts described by quantum mechanics. But there's a widespread consensus that the thermodynamical facts are in some sense autonomous (Sklar, 1993, p. 344).[2] The project of this paper is to articulate the sense in which the facts described by thermodynamics and other special sciences are autonomous. Moreover, I want to say that this autonomy is not a feature of our perspective or our cognitive limitations; we don't practice biology and chemistry because quantum physics is too hard for us. There is a genuine, or worldly, feature that can explain the methodological independence. Physics *really* doesn't matter.

Nevertheless the view that physics is totally irrelevant for higher-level sciences such as biology is implausible; the prevalence of MRI scans (which use nuclear magnetic resonance) for the detection of disease, positron emission tomography (PET) scans and X-ray crystallography are just a few examples of how the higher-level sciences are not *completely* insulated from more fundamental levels, such as physics.

Whether there are 'new' metaphysical ingredients at higher levels and how the special sciences are related to more fundamental sciences —whether they are reducible, emergent, or completely disunified— is deeply controversial (cf. Nagel (1961); McLaughlin (1992); Cartwright (1999)). But it is uncontroversial that the special sciences are (at least) *methodologically* autonomous of fundamental physics. A theory of quantum gravity is highly sought after — but not because any psychologist believes it will help with their research on implicit bias. The practice and institutional structure of

[1] List (2019) claims talk of levels is 'ubiquitous', and quotes Kim (2000, p. 16), who says "[t]alk of levels of organization, descriptions or languages, of analysis, of explanation, and the like is encountered everywhere".

[2] "Thermodynamics is silent about the nature of matter. In particular it does not insist that matter is made of particles, and in fact as long as one works within thermodynamics (and does not proceed to statistical mechanics) one may assume, for all practical purposes, that matter is continuous. In this sense thermodynamics is autonomous of mechanics" (Hemmo and Shenker, 2012, p. 27).

scientific inquiry is such that psychologists and biologists need not learn much, or indeed anything, about fundamental physics in order to make progress in their fields. Thus, this methodological autonomy is a procedural independence of the distinct disciplines. Put another way: even the most ardent reductionist would not require that all scientific inquiry other than fundamental physics be stopped. We learn and gather scientific evidence in a way that is methodologically independent of fundamental physics.

How should we explain this methodological independence in light of the relevance of physics? If the Schrödinger equation applies to wildebeest migration, why does this not matter for ecologists' methodology? In other words: why doesn't physics matter more? This is the explanatory puzzle at the heart of this paper. The solution will be that the special sciences are methodologically independent of physics because the facts uncovered by the special sciences are autonomous — according to my account of autonomy.

In section 2, I discuss Fodor's account of autonomy, and its shortcomings. In section 3, I give an illustration of the explanatory puzzle, and show how conceptualising autonomy as either 'full' or 'apparent' makes the puzzle worse; either you can't do justice to the relevance of physics, or you can't do justice to the independence from physics. In section 4 I propose a solution to this, through my account of 'generalised autonomy'. Generalised autonomy is compatible with both reduction and non-reduction. This makes my account a striking departure from Fodor's, and I discuss this difference in section 5. Then, in sections 6, 7 and 8 I connect generalised autonomy to dynamical, causal and nomic autonomy respectively. In section 9, I discuss how —sometimes— we can specify *which* lower-level details don't matter, and make a further distinction between types of autonomy. In section 10 I show how generalised autonomy helps with the explanatory puzzle. I argue that generalised autonomy not only captures the senses in which physics is both relevant and irrelevant to the special science facts, but it also can account for the epistemic role played by non-fundamental theories in the search for more fundamental theories. The special sciences wouldn't be able to do this if the special science facts were fully autonomous rather than partially autonomous in the manner of 'generalised autonomy'. Section 11 concludes.

2 Fodor's view of autonomy

The view that the special sciences are autonomous originates in the contemporary literature with Fodor (1974). Fodor discusses the autonomy of the special sciences in service of his non-reductive physicalism, according to which special science facts supervene on, but are not reducible to, the fundamental physical facts. Part and parcel of this view is Fodor's realism about special science kinds, claiming that they exist 'over and above' the fundamental physical kinds. Loewer (2009) criticises Fodor's position for wavering between two readings of autonomy:

1. The first reading of autonomy is a methodological autonomy. Yet methodolog-

ical autonomy — the independence of ecologists from quantum physicists — is seemingly a sociological fact about the structure of scientific practice, and therefore cannot licence metaphysical conclusions (without substantive further assumptions[3]). After all, this methodological autonomy could just be an unfortunate consequence of our cognitive limitations: we have to just get on with doing ecology paying no attention to quantum mechanics because the Schrödinger equation for such large systems is just too fiendish *for us* to solve. Such an attitude lends itself to an instrumentalism about the special sciences which is diametrically opposed to Fodor's view, but still compatible with the methodological independence of the special sciences. But leaving aside Fodor's project, methodological autonomy is precisely what we want to explain here, and it cannot be both explanans and explanandum.

2. The second reading is a metaphysical or 'full' autonomy. Loewer illustrates this type of autonomy with a popular metaphor. The idea is that whilst the fundamental level has been fixed, God is free to pick the way the world could be according to the special sciences: there could be square oranges, a different menagerie of animals and entirely different economies. But Loewer argues that this metaphysical autonomy is implausible given Fodor's commitment to supervenience physicalism. According to supervenience physicalism, there can be no change at the special science level without a change at the (fundamental) physical level, and so once God has fixed this fundamental level, there is no 'choice' of how things can be at the special science level. Hence, Loewer concludes that the special sciences kinds in Fodor's ontology are mere idle cogs.

My project in this paper is not to defend Fodor. In addition to Loewer's objections, Fodor has been criticised for arguing unsuccessfully for the irreducibility of the special sciences. (For example, Sober (1999) asks: why should a smattering of the word 'or' prevent reduction?). And the combination of non-reduction and supervenience is an unstable balancing point between the primacy of physics and the importance of the special sciences — when Beth's theorem applies, supervenience collapses into reduction, cf. Hellman and Thompson (1976); Butterfield (2011a).

Supervenience is nonetheless a useful tool for spelling out the sense in which physics, or other lower-level facts, are relevant to the special sciences or higher-level facts. Here we need not endorse a supervenience physicalism that claims that there is a single fundamental level upon which all else supervenes, since the situation could be more piecemeal: for any higher-level fact, there just needs to be *some* lower-level fact that subvenes it. There are many further metaphysical and explanatory questions about the relationship between levels, but supervenience suffices to capture the aspects required for our project.[4] Supervenience has fallen out of vogue in recent years,

[3]The scientific realist might argue that the empirical success of the special sciences is sufficient to commit to the ontology of the special sciences, but the autonomy — methodological or otherwise — doesn't come into this argument.

[4]As such, my account should be compatible with a variety of metaphysical explanations that go beyond supervenience.

largely because it is thought to be too weak to be informative, especially in the philosophy of mind (Heil, 1998). But weakness is a virtue when employing assumptions, which is how I am using supervenience here. Moreover, the assumption that supervenience holds between levels is not a vague hope; a very precise account of how levels are related by supervenience maps has been given by List (2019). Supervenience is usually taken to be a relation between facts and, in keeping with this, List's levels are not the old-fashioned mereological levels of Oppenheim and Putnam (1958) criticised by, *inter alia*, Potochnik (2017). On List's framework, there is a range of different types of levels: ontological levels, levels of grain, levels of dynamics and levels of description. The latter correspond to the different languages of different scientific theories; for example, a higher-level theory might talk about a small number of macrovariables, whereas the lower-level theory might employ (a much larger number of) microvariables. As such, we can talk of macrovariables supervening upon microvariables; the value of micro or macrovariable that is instantiated is just a micro or macrofact.

In general, supervenience constrains the relationship between levels, and here we use this relationship to express the sense in which the lower level is 'relevant' to the higher level. According to this supervenience view of levels and relevance, 'a lower-level fact is relevant for a higher-level fact' is cashed out as there can be no change to the higher level without a change to the lower level. This is a strong sense of 'relevance', as the lower level is so relevant as to determine the higher level.[5] Having this strong sense of relevance makes the explanatory puzzle harder, and so a worthier challenge to solve.

Fodor's account of autonomy does not serve his aims (according to Loewer), and it does not serve our explanatory project. The methodological reading is precisely what we want to explain, and the 'full' autonomy reading is unnecessarily strong. Autonomy can be partial. To take a political example, a region can be autonomous of a nation state: it is self-governing or independent in various ways. Yet it will depend on the state in various ways — perhaps financially, or in terms of trade agreements — and not least for recognition as an autonomous region. The more facts 'Y' with respect to which another fact 'X' is autonomous the *greater* the autonomy of X. In the limiting case where Y is 'everything else', then we might say X is 'fully autonomous'. Failing to recognise that autonomy can be partial makes the explanatory puzzle more vexing — as we will see in the next section.

3 The explanatory puzzle

The puzzle: if physics is relevant for the special sciences, in the sense that the physical facts subvene the special science facts, why doesn't physics matter more to these sciences? How can we explain the methodological independence of the special sciences from physics?

[5]Of course, not everyone understands supervenience as determination. But the key point is that the worry from the philosophy of mind context that supervenience is too *weak* to be useful does not carry across to this context.

Here 'physics' refers solely to fundamental physics, because many theories within physics are non-fundamental, and so are special sciences.[6] Thermodynamics is a case of a special science within physics and I will use it as an example throughout. Furthermore, many have remarked on thermodynamics as an 'autonomous science' (Sklar, 1993, p. 344). The idea originates with Duhem, Mach and the energetists who "insisted that these [thermodynamic] principles were *autonomous* phenomenological laws that needed no further grounding in some other physical principles" (Sklar, 2015, S1).

But we needn't be focused solely on the fundamental level, and we needn't follow Fodor in endorsing supervenience physicalism in order to get the interesting question off the ground. All that's required is that there is a *relative* fundamentality ordering: some special science facts have at least one lower-level fact underlying them. As such, there needn't exist one *single* physical level onto which all other levels supervene. There might be a chemical fact underlying a biological fact, or a neurophysiological fact underlying an economic fact. On the relative fundamentality view, the question is then not only, Why doesn't fundamental physics matter more, but rather, Why don't the lower-level details matter more?

Here is an illustration of the puzzle. The quantum nature of matter is crucial to its stability, and so there is a sense in which the regularities of thermodynamics are not autonomous. Had matter been classical, the atom would not have been stable, and there would be no actual macrofacts for thermodynamics to describe. Indeed, this is a very general point: consider how much of science is clearly unperturbed by the discovery that matter is quantum, not classical. 1905 heralded a revolution in physics, but not in ecology, geology or chemistry. But nonetheless these sciences are not autonomous of the quantum nature of matter: for if matter were classical rather than quantum, it wouldn't be stable and so we wouldn't have the macrofacts, laws and causes that those special sciences describe.

More speculatively, not only is the quantum nature of matter crucial to its stability, quantum mechanics is required for nature to process information in a digital rather than analogue format in order to avoid degradation. The discrete nature of energy levels in quantum mechanics allows for better storage of information — and, as a consequence, Schrödinger (1967) argues that quantum mechanics is crucial to the possibility of life itself.

It seems that the fact that nature is quantum rather than classical is relevant for the special sciences, yet the revolution in physics does not have much impact in other sciences. Whilst there are some explanations which depend explicitly on quantum mechanics which I will touch upon later, for the most part biologists can ignore quantum mechanics; the practice of biology was not revolutionised by the discovery of quantum mechanics. This leads to the puzzle: why are the special sciences methodologically autonomous, especially since it seems quantum mechanics is relevant?

The answer is that methodological autonomy is not the only form of autonomy possessed by the special sciences. The facts — the laws, causal claims and dynamics

[6] Special sciences are just non-fundamental theories.

— discovered by the special sciences are in some way autonomous of physics. But to flesh out this answer we need an account of autonomy, and the literature promulgates a false dichotomy which prevents any satisfactory resolution of the puzzle. This false dichotomy is between full autonomy and merely apparent autonomy.

According to full autonomy, the physical facts do not matter at all. The methodological independence of the special sciences is explained by the full autonomy of the special science facts; physics would be completely irrelevant, since God could have chosen different biological facts even after decreeing all the physical facts. Yet this flies in the face of the assumption of supervenience, the relevance of physics in various instruments such as microscopes and — more prosaically — the relevance of physics for the technology that surrounds us daily.[7]

On the other hand, anything less than the complete or full autonomy demonstrated by 'choosing the higher level any which way' is deemed to be mere apparent autonomy. Loewer's 'metaphysical autonomy' is an example of this full autonomy. Batterman discusses how a reductionist physicist "challenges the idea that phenomenological theories like continuum mechanics, or more generally special sciences like psychology and economics exhibit *genuine* autonomy from the more fundamental 'structural' theories of high energy physics. If all generalizations of such phenomenological theories can really be described and predicted perfectly well by structural theories, then their *apparent* autonomy is just that — apparent" (Batterman, 2018, p. 859), emphasis added.

If examples of autonomous facts must either be fully autonomous or merely apparently autonomous, the puzzle deepens. On the assumption that a higher-level fact supervenes on *some* lower-level fact, any autonomy does not qualify as full autonomy and so must be merely apparent. That is, the illustrative example of quantum mechanics shows that the lower-level fact and its details really *do* matter, it just *appears* that they don't. If the autonomy of biological facts is just a trick of the light, a lucky mirage, it seems like a very pervasive one. At this point, one feels driven towards accepting the explanation that methodological autonomy is just an inevitable consequence of our cognitive limitations. Biologists have to proceed as if physics doesn't matter, despite the fact that really it does.

Citing our cognitive limitations to explain methodological autonomy leads to an instrumentalist view of the special sciences; they are just useful tools for *us*. An argument for this instrumentalist view goes as follows. From the comfort of our armchairs it seems astonishing that nature would be so kind as to stratify the world into levels or layers that we can investigate (fairly) independently of one another. Yet being able to cleave science into institutionally distinct disciplines is key to making progress: if we had to consult the standard model of particle physics every time we wanted to understand tension in bridges or osmosis across cell membranes, progress in gathering scientific knowledge would stall. Since this methodological autonomy is so central to

[7]Proponents of disunity do not have a problem explaining the independence of the special sciences; it is instead the relevance of physics for our everyday lives and other sciences (such as the use of nuclear magnetic imaging) that is a puzzle for them.

our scientific endeavours — perhaps even a necessary condition of scientific methodology — one might worry that these divisions are just something that *we*, with our limited cognitive powers, have added to the picture — rather than being a feature of Nature herself.

But this view just entrenches the puzzle. If the methodological autonomy of the special sciences is forced on us just because *we* cannot use the more fundamental theory to describe the phenomena of the special sciences, how can we explain the empirical successes of the special sciences? It seems like a stroke of luck that our cognitive limitations mesh so successfully with the world that ignoring physics hasn't impeded the success of the special sciences.

Thus, the current accounts of autonomy of the special science facts cannot explain the methodological independence of the special sciences — largely because of the false dichotomy between full autonomy and mere apparent autonomy. In the next section, I develop a new account of autonomy. This concept of autonomy captures the idea that some of the lower-level details genuinely do not matter for the special science, or higher-level, description. The autonomy is worldly, rather than just apparant to us. But this concept of autonomy is partial: some lower-level details *do* matter. Autonomy is not all or nothing. This view will allow us to explain methodological autonomy without making our cognitive limitations centre stage.

4 Developing an account of autonomy: generalised autonomy

Earlier I said that the higher-level (or special science) facts supervene on the lower-level (or physical) facts. Alternatively, we could talk of microfacts and macrofacts.[8] Though ultimately supervenience characterises relations between facts, it sometimes helps to be a bit more fine-grained: we can talk about macrostates (the specification of a set of macrovariables) and microstates (the specification of a set of microvariables). We will also be interested in the dependencies between these states. The macrostates, A and B, are often — but not always[9] — temporally ordered, and the macrodependencies between them could be causal (B causes A), nomic (B is connected to A by a law) or dynamical (state B is taken to A by the dynamics).

My account of autonomy needn't specify what type of macrodependency 'B-A' is. But once we specify the type of macrodependency present in a given example we can make connections to the respective types of autonomy; dynamical autonomy, nomic autonomy and causal autonomy. I have therefore called this account 'generalised autonomy', since it unifies the definitions of dynamical autonomy in differential equations (Robinson, 2004), causal autonomy in an interventionist framework (Woodward, 2018), and nomic autonomy (Franklin, 2020).

[8]For brevity I will often use this locution, but I do not intend to imply that levels are necessarily scale-dependent (cf. Potochnik (2017)); the reader can substitute 'micro' for 'lower-level' throughout, if preferred.

[9]See the ideal gas law discussed later on, for example.

Here is the key thesis of generalised autonomy: a microfact b is unconditionally relevant to the macrofact A, but conditional on macrofact B, b is irrelevant to A. One way of putting this: the macrodependence between A and B *screens off* the microdetails.

In section 4.1 I explicate this key idea, which is is inspired by, and stems from, Woodward's account of explanatory autonomy (Woodward, 2018). Then, in section 4.2, I discuss how the 'screening off' is sometimes only approximate, and how this approximation can be made more accurate using a strategy from the *ceteris paribus* law literature. Section 4.3 explains how this account is one of partial autonomy, unlike Fodor's account, and in section 5 I discuss another difference from Fodor's account: generalised autonomy is compatible with reduction. Then in sections 6, 7 and 8 I discuss the types of autonomy unified by my account.

4.1 What is generalised autonomy?

Under normal conditions, the water in your kettle boils at $100°C$. There is a macrodependency between one macrofact 'B' (temperature) and another 'A' (boiling). This macrodependency describes what goes on in your kettle when you want a cup of tea, but there are other relevant facts; in particular there are lower-level facts upon which B and A supervene. In this case, the molecular motion of the water molecules is very relevant to whether your kettle boils: the faster the velocity, the more likely it is that kettle will boil. The particular motion of the water molecules in the kettle is a lower-level, or 'microlevel', fact, which I will call a 'microfact'. As discussed earlier, sometimes this microfact can be called a 'microstate'. Here I denote such a microfact as 'b' and take b to be the supervenience basis of B.

The dependency between b and B is captured by the following 'unconditional relevance condition', and, in general, these (in)dependencies can be represented by probabilities. Much more could be said about the nature of the probabilities in question. In some cases, the objective probabilities could be fundamental chances in an indeterministic world (Schaffer (2007); Lewis (1986)). Alternatively, the objective probabilities could be deterministic chances (List and Pivato, 2015; Hoefer, 2007; Handfield and Wilson, 2014). But my argument is also applicable to an epistemic reading, so long as the probabilities are not purely subjective credences but are in a suitable sense epistemically normative.[10] However, for our purposes here, many of the details can be left aside, since no probabilities need be explicitly calculated nor ascribed numerical values.

- **Unconditional relevance:** conditional on a particular microfact b, the probability of the macrofact A obtaining increases: $P(A|b) > P(A)$. Under certain

[10]We can capture the required objectivity of probability by requiring that something like Lewis' principal principle holds: rational credences must be aligned with objective probabilities. As such, however we choose to understand the metaphysics of probability, the key requirement is that probability as used in this paper must be rationally guiding in the way prescribed by whatever probability coordination principle we in fact accept, if not the principal principle.

circumstances, $P(A|b) = 1$.[11]

Much of the fine-grained detail of a given microfact doesn't matter, because it doesn't influence A. This means that there are other microfacts b_1, b_2... b_N for which $P(A|b_i) > P(A)$. To return to the temperature example, there is a range of molecular motions that lead to the kettle boiling; exchanging the molecule 320510 with molecule 105503 makes no macro difference. We can think of these different microfacts as forming an equivalence class, that is, a collection of microstates which influence macrofact A. (For now, this equivalence class is a purely formal device, but as I will describe in later sections, we can understand these equivalence classes as partitions over the state space of a dynamical theory.)

Now let's assume that *which* member of this equivalence class of microfacts is instantiated doesn't matter for the occurrence of A (we will relax this assumption in the next section). Another way of showing that some lower-level feature is irrelevant, is by comparing two microfacts which differ with respect to that feature: $P(A|B\&b_1) = P(A|B\&b_2)$.

This already indicates that some of the microdetails will be irrelevant, since differences between the members of the equivalence class don't matter. Once we stipulate the macrofact that this equivalence class of microfacts subvenes, the lower-level details don't matter. In the boiling example, given the temperature of the water is $100°C$ (macrostate B), the molecular motion is irrelevant.

So far we have seen that the molecular motion is unconditionally relevant for the boiling, because the boiling depends on the molecular motion. But once we conditionalise on the macrofact B (the temperature is 100°C), any further fine-grained details about the motion of the molecules are irrelevant. This is *conditional irrelevance*, and is a form of screening off. For now, we assume that the B-A is a strict (though not necessarily deterministic) dependency, and conditional irrelevance is defined as follows.

- **Conditional irrelevance (strict version):** $P(A|B\&b) = P(A|B) = x$, where $0 \leq x \leq 1$.

This doesn't require that the macrodependency B-A is deterministic ($P(A|B) = 1$), because if $x < 1$, B could still exactly screen off b.

Because of the macrodependency B-A, conditionalising on B screens off the influence of b on A. In other words, all the influence of the underlying microfact is mediated through B. This is particularly clear if we represent the relations graphically. Moreover, this shows how the conditional irrelevance condition is akin to a Markov condition in a multi-level setting.

In the Figure below, the left hand diagram shows a directed acyclic graph, where the horizontal arrows represent causal dependence. The causal Markov condition requires that the nearest ancestors, i.e. most proximate causes, screen off more distant ancestors, so that Y screens off X from Z. The middle diagram is analogous to

[11] $P(A|b) = 1$ if the microdependencies take every member of the supervenience basis of B to the supervenience basis of A.

the putative grounding cases discussed by Schaffer (2016); the vertical arrows represent grounding relations. There is a grounding Markov condition: the psychological facts (P) screen off the neurological facts (N) from the well-being facts (W). Stern and Eva (2020) generalise the interventionist framework to a multilevel setting, where the vertical arrows represent supervenience relations, and the horizontal arrows represent causal relations (although for our purposes, the nature of the macrodependency needn't be causal). The multi-level Markov condition is represented in the right hand diagram: B screens off b from A.

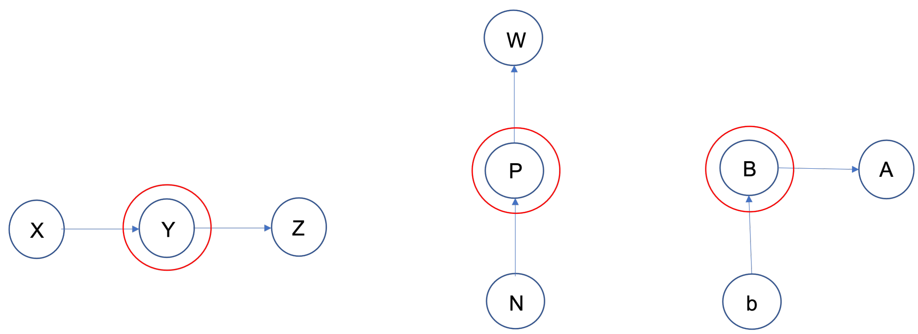

Figure 1: In each of these diagrams (causal, grounding or multi-dimensional respectively), the circled variable (or fact) screens off the less proximate ancestors (X, N and b respectively) from Z, W and A facts respectively. As such, each figure demonstrates a Markov condition.

> **Generalised Autonomy (GA)**: a macrofact A has generalised autonomy if the microfact b is unconditionally relevant, but conditional on macrofact B the microfact b is irrelevant. Here autonomy is a three-relation: A is autonomous of b given B.

Here are two brief examples of generalised autonomy:

1. The Game of Life demonstrates generalised autonomy. There is a menagerie of different entities in the Game of Life, such as the gliders shown in Figure 2. There are laws, or macrodynamics, such as 'gliders travel at a quarter of the maximum speed in a diagonal direction'. These glider dynamics determine the later position (A) from the earlier position (B) of the glider. But the microstate of the grid and the grid microdynamics are unconditionally relevant to the later macrostate A (the later position of the glider). Nonetheless conditional on the earlier macrostate B (the earlier position of the glider), the fine-grained details of the microstate are irrelevant because they are screened off by the glider dynamics.

2. Generalised autonomy is exhibited in the general, albeit less interesting, example of variables with different grain. To take a typical example, the exact mass

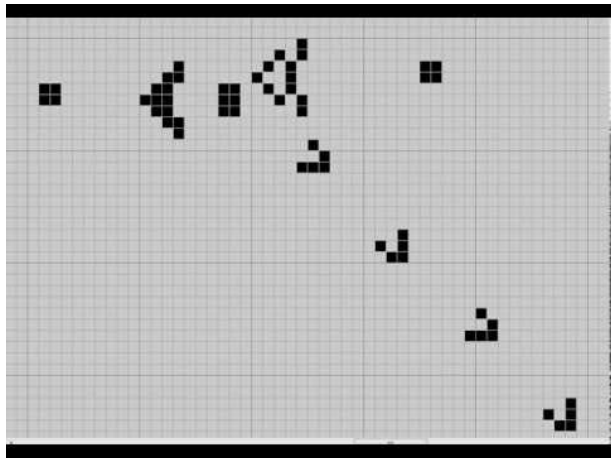

Figure 2: A snapshot of the Game of Life grid displaying gliders travelling down to the bottom right hand corner of the grid.

of the rock is unconditionally relevant to the window breaking but once we conditionalise on the mass being over 5kg (as a putative threshold for the glass to smash), the exact mass is irrelevant to the fact the window broke.[12] A more interesting case is the exact (analogue) shape of a waveform carrying a digital data stream vs. the digital data encoded. For the information to get transmitted as intended, it matters that the datastream be the right shape within broad parameters (which we can assume the source guarantees), but it doesn't matter which of the numerous suitable exact waveforms was used. Here it's clear what is getting screened off: the exact shape of the waveform is unconditionally relevant but conditionally irrelevant.

4.2 Deviant microstates, *ceteris paribus* laws and background conditions

So far, we have assumed that the macrodependency is strict, and — relatedly — that it does not which member of the equivalence class (of microfacts $b_1, b_2 ... b_N$ subvening the macrofact B) is instantiated. Now we relax this assumption, and expand our attention of the supervenience basis of B to include that of A. This brings into consideration the *micro*dependencies thus far set aside. We need to consider this because, as emphasised by Fodor, one type of macrodependency — special science laws — is usually thought to have exceptions, and so are called *ceteris paribus* laws. For example, special science laws like Mendel's law, Snell's law and the law of entropy increase are often understood as *ceteris paribus* laws.

Here it helps to start with a case of levels of dynamics, as discussed by Fodor, and as I will analyse further in section 6. Figure 3 differs from Figure 1 in that the horizon-

[12]Of course, the exact mass might not be irrelevant to the window breaking in the exact way that it did.

tal arrows represent dynamics rather than causal dependence, and the supervenience basis of A and the microdynamics T are included.

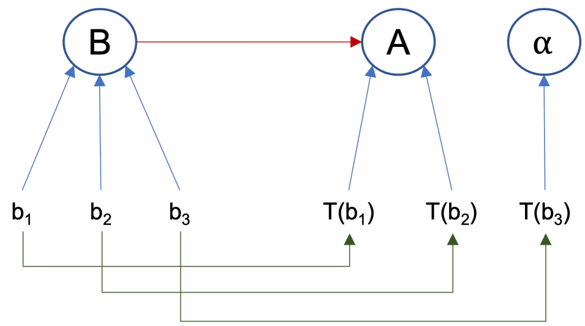

Figure 3: Here not every member of the supervenience basis of B is taken by the microdynamics T to a member of the supervenience basis of A. Instead b_3 is part of the supervenience basis of a distinct macrofact α. In statistical mechanics, such a microstate is sometimes called a 'deviant microstate'.

When the microdynamics T take every microstate in the supervenience basis of B to the supervenience basis of A, we have a strict macrodependency, and B screens off b exactly. All of the influence of b is mediated through B.

Exceptions to the macrodependency B-A occur when one (or more) member(s) of the supervenience basis of B ($b_i ... b_N$) is not taken by the microdynamics to the supervenience basis of A, but instead is a member of the supervenience basis α, as shown in Figure 3. And where there is a deviant microstate, we only have approximate screening off. In contrast to the strict case, it *does* matter which member of the equivalence class is instantiated. In the strict case, $P(A|b_1) = P(A|b_N)$ but in the deviant microstate case, $P(A|b_1) \neq P(A|b_3)$.

The Game of Life example discussed in the previous section is analogous to the statistical mechanical case study we will consider in section 6.1. The macrofacts about gliders have generalised autonomy. However, the glider dynamics macrofacts only hold until a particularly pernicious piece of debris floats by and destroys the glider. So it is only over certain timescales, and for particular initial microstates (since these determine whether the conditions are right for gliders) that the glider dynamics screen off the full microstate. There are some microstates (i.e. configurations of the grid) that subvene a particular glider position that do not lead to the glider being in the later position predicted by the glider dynamics, because the piece of debris has destroyed the glider pattern: these are the 'deviant microstates'.

In general, screening off is approximate, since there are some initial microstates for $P(A|B\&b_3) \neq P(A|B)$. So in full generality,

- **Conditional irrelevance:** $P(A|B\&b) \approx P(A|B) = x$ where $0 \leq x \leq 1$.

What is the significance of the conditional irrelevance condition only being approximate? Should we worry that this makes generalised autonomy mere *apparent* autonomy? No. Deviant microstates emphasise the way in which the microfacts matter, and thus highlight that generalised autonomy is not 'full autonomy'. However, this doesn't mean the autonomy of the macrofact is an illusion, as I argued earlier. Moreover, we've already explicitly admitted that 'generalised autonomy' is not full autonomy with the unconditional relevance condition. The approximate screening off tells us that conditional on macrofact B, the microfacts don't matter — most of the time. The autonomy hasn't disappeared. We can increase our precision, even sometimes achieve exact screening off, by drawing on strategies used in both *ceteris paribus* laws literature and the philosophy of statistical mechanics.

Here's how: we stipulate background conditions that rule out the exceptions or deviant microstates. These background conditions specify the timescales, situations or contexts in which we can expect a given macrodependency to be successful. Snell's law is a *ceteris paribus* law, because it is not universal; it only applies to isotropic media (Cartwright, 1983). The law of demand requires the caveat (or background conditions) that we are operating under conditions of ideal competition, and that other markets and variables make negligible difference. To take an imaginary example, imagine that there is a particular gene that protects against the types of cancer caused by smoking. The physical microstate that describes a smoker with this gene—a deviant microstate—is not screened off. However, if we put this exception in the background conditions (under normal background conditions, i.e. those without this special gene), then we have exact screening off.

In a way, this background conditions stipulation isn't really adding anything new to the account, since we normally need to include some kind of background condition. For example, in the kettle boiling example, we had to state 'under usual background conditions'; if you were trying to boil your kettle on a commercial airplane or on top of Everest, the temperature at which it boils would be lower. In the *ceteris paribus* laws literature adding in such caveats known as adding in 'completers' (Reutlinger et al., 2020). Whether this must be done in practice or is just a promissory note depends on one's view of *ceteris paribus* laws, which in turn depends on one's metaphysics of laws (cf. Earman and Roberts (1999)). However, it is worth pointing out one connection to the philosophy of statistical mechanics. Here the deviant microstates can be ruled out by a background condition, which stipulates an initial microstate is 'Natural' or simple in a particular way (Wallace, 2019).

How important is it that we state these background conditions and improve the approximation? In practice, like all approximations, whether we *want* to improve the approximation will depend on our goals and purposes. Newtonian mechanics is a good enough approximation for calculating the trajectory of a football or sending astronauts to the moon, but it isn't a good enough approximation for GPS. If you are concerned with health policy, then the macrodependency 'smoking causes cancer' is fine-grained enough. But if you are interested in personalised, or individualised medicine, then this might not be fine-grained enough. With the individualised goals in mind, we would want to finesse the macrodependency by adding other background conditions, such as your genetic predisposition, how many cigarettes you smoke, other comorbidities. But even with the finest-grained goals, i.e. the most highly specified background conditions, and even if the macrodependency is not rendered deterministic, there may still be exact screening off. Not every smoker with all the particular risk factors will develop cancer: $P(A|B) < 1$. There can be exceptions, but these exceptions might be truly random (this is made particularly plausible when one considers that the mechanism behind cell mutation is a truly random process). When the microstates that are 'deviant' are truly random, the macrodependency will be of the form $P(A|B) = x$, where $x < 1$, and the screening off will be exact: $P(A|b\&B) = x$ (rather than $\approx x$).

Of course, sometimes we won't be in the epistemic position to specify the requisite background conditions to restore exact screening off. But knowing that they could in principle be discovered reassures us that the approximate nature of the screening off is a mere epistemic problem, and needn't detract from the conditional irrelevance condition. But sometimes, especially if the approximation is not good enough for our purposes, this might force us to find out when and where the macrodependency is unsuccessful. Here details from the lower-level theory could be helpful for improving the higher-level description, and this brings us to a consideration the partial nature of generalised autonomy.

4.3 Dimensions of the partial nature of generalised autonomy

As we've already seen, generalised autonomy is a partial form of autonomy. My account embraces the idea that the autonomy of the special sciences is not full autonomy by explicitly modelling the dependence of the higher level on the lower level: the unconditional relevance condition. This means that the microfacts are never wholly irrelevant and the macrofact does not float entirely free of the lower level. In other words, the macrofact is not compatible with *all* and *any* lower-level fact: instead we require that the microfact b must be 'compatible with' i.e. subvene macrofact B.

The second sense in which generalised autonomy is partial is evident when considering 'deviant' microstates which are the basis of the exceptions to special science laws familiar from the *ceteris paribus* laws literature. In the case of the deviant microstates, the conditional irrelevance condition may be only an approximate screening off — unless we can find the right background conditions to restore the exact screening off. Discovering these background conditions might involve detailed consideration of the microfacts, and might reveal the autonomy to be limited to certain timescales or to certain degrees of approximation. To return to our earlier illustration: macrofacts about glider positions have generalised autonomy — until a pernicious piece of debris floats by and destroys the set up. Whether the piece of debris will do this, and when, is determined by the initial microstate of the grid. This means that the lower level might set up the conditions or circumstances for the macrodependency to be successful. This is, I believe, what Loewer (2008) means when he says that physics itself is responsible for why there is anything other than physics.

There is one final sense in which generalised autonomy is partial, which I have not yet discussed. Generalised autonomy applies to a given macrofact, but not necessarily to a whole *theory*. In particular, there could be a set of related macrofacts which are not autonomous. For example, the migration of the European robin can be explained by evolutionary biology, and this macrofact is autonomous of the underlying microstate. But there is a related macrofact that is not autonomous. The mechanism of how the robin navigates requires a quantum mechanical description (Rodgers, 2009; Shrapnel, 2014). As such, not all facts about avian migration are autonomous of the more fundamental descriptions, although some are. Particular macrofacts might be autonomous, but whole theories, or levels — as collections of macrofacts — might contain some macrofacts that are not. Facts, rather than theories, are what can be autonomous. Nonetheless, we can understand claims that a theory, such as thermodynamics, is an autonomous special science as saying that some of the central macrofacts have generalised autonomy.

Next, we consider the connection between generalised autonomy and the status of particular scientific theories — namely, whether they are reducible or not.

5 Generalised autonomy is independent of reduction

In the classic debate about reduction between Oppenheim and Putnam (1958) and Fodor (1974), Fodor's position that the special sciences are autonomous is tied to the irreducibility of the special sciences. But, contra Fodor, generalised autonomy is compatible with — but doesn't require — inter-theoretic reduction. For instance, in the boiling kettle example, the higher-level variables (e.g. temperature) and their interrelations can be understood through (or derived from) the lower-level facts about molecular motion. Likewise, in the case study of statistical mechanical macrodynamics, we will see that these macrodynamics can be constructed from the underlying microdynamics.

Of course, we could quibble about whether these are really complete and successful reductions, but since it is at the very least plausible that they are indeed cases of reduction, generalised autonomy does not hang on anti-reduction, unlike Fodor's account. The macrodependencies screen off the microdetails, and this allows the macrofacts to be autonomous. This is compatible with giving a bottom-up explanation of the macrofacts, and so generalised autonomy is compatible with reduction.

Equally, generalised autonomy doesn't require reduction. We do not need to specify *which* details don't matter, i.e. those that are conditionally irrelevant. There is no general feature about lower-level details; instead it will largely be a case- and discipline-specific matter to determine which details don't matter, as we will see in the case studies below. Having a successful inter-theoretic reduction would help establish which details do matter. Requiring these details might inadvertently lead to requiring inter-theoretic reduction as part of generalised autonomy. But since we do not need to specify *which* details don't matter in order to judge a macrofact to have generalised autonomy, we don't require inter-theoretic reduction, and so leave open the possibility of non-reduction. For instance, whilst it is uncontested that in practice we currently have no way to reduce the biological description of wildebeest migration to the level of quantum mechanics, it is contested whether *in principle* there is a potential reduction. (Nonetheless here I assume that the Schrödinger equation applies to the wildebeest, but say nothing about whether or not a reduction can be achieved since there's no uncontested connection or bridge between the quantum mechanical description and the biological description). Regardless of this contested corner of the reduction debate, my account allows us to express our confidence that the wildebeest case is autonomous, without needing to specify or know which parts of the quantum mechanical description are relevant and which are not.

Generalised autonomy's compatibility with both reduction and non-reduction leads to a second divergence from the Fodor-Putnam debate. A feature of that debate was the historical contingency of reduction and autonomy: for Putnam it is an empirical question whether reductionism succeeds. Likewise, Fodor claims that *thus far*, i.e. 'after all these years', the special sciences are still autonomous — but nonetheless the threat of reduction, and consequent loss of autonomy still looms.

In contrast, on my view, reduction does not threaten whether a macrofact has generalised autonomy. A fully explicated reduction might allow us to say which details matter and which don't, and might shed light on the timescales or domains in which the higher-level theory is a successful one — but it doesn't threaten the autonomy of the macrofacts described by that theory. In fact, the converse is true: a fully worked-out reduction might allow us to fill in the requisite background conditions to remove any element of approximation in the conditional irrelevance condition.

In part, generalised autonomy is orthogonal to reduction because inter-theoretic reduction is an epistemological relation: whether it holds depends on whether or not this body of knowledge, or theory, can be suitably connected or derived from another body of knowledge. In contrast, generalised autonomy is a worldly feature: whether it holds depends on whether a given macrodependency screens off a particular microfact — and this does not depend on the contingency of our current knowledge. Since this screening off remains even where there is inter-theoretic reduction, we can explain why we don't just always use the reducing lower-level theory — some of the details are truly irrelevant to certain macrofacts and so their inclusion wouldn't count as an improvement over the special science theory.

Having explicated the core features of generalised autonomy, in the next three sections I show how generalised autonomy is connected to dynamical, causal and nomic autonomy respectively.

6 Dynamical autonomy

In this section I discuss 'dynamical autonomy' as it has been considered in the literature, and show how — once one element is added — this is a form of generalised autonomy.

A dynamical theory has two components: (i) the kinematics, which describe possible states of the system, i.e. the state-space, and (ii) the dynamics, which dictates how the system's state changes over time as it moves through the space of possible states. For example, the system's state in phase space represents its position and momenta, and the Hamiltonian dynamics specify how these properties change over time. Differential equations often provide a natural way of representing these dynamics.[13]

The theory of differential equations provides a definition of autonomy as follows.[14] A differential equation for the evolution of the variable y is autonomous of some other variable x if neither x nor t appear explicitly in the equation, as in $\frac{dy}{dt} = f(y)$. If x does not explicitly appear, then the dynamical evolution of y is independent of whatever values x takes (Robinson, 2004). The condition that the evolution does not depend on t — that t does not appear on the right hand side of the equation — means that the way y evolves can't change (e.g. from a sine wave to a sawtooth when $t > 10$). This ensures that the evolution of y does not depend implicitly on x: it rules out any 'covert' dependence on x (i.e. $t > 10$ might be implicitly encoding that x is greater than a certain value).

Of course, x might not appear in the equation for the temporal evolution of y because x is unconditionally *ir*relevant for y. The variable for the trajectory (y_{park}) of the football I have just kicked across the park does not depend on the location of the spare football (x_{home}) I left at home; the spare football is unconditionally irrelevant for the dynamical evolution of the football in the park. This type of dynamical autonomy is so dull, it is usually ignored. But when x is not wholly irrelevant for y, then things become more interesting. For example, y might be the centre of mass of the park football (an infinitesimally small point-like object in space) and x might be not just one variable but an exceedingly large number of variables representing the position of each molecule in the football. Clearly x is relevant for y, since x subvenes y, and some might go as far as saying that x *determines* y. Nonetheless y (the centre of mass of the football) is dynamically autonomous of x (the position of every molecule in the football): a differential equation can be given for y (for example, the differential equations would be the familiar Newtonian equations if my park is on Earth, and if I'm incapable of kicking the football to relativistic velocities) — and these equations makes no reference to the molecules in the ball.

Once x is a variable that is unconditionally relevant for y — as is achieved by claiming y supervenes on x — then dynamical autonomy becomes far more interesting. Furthermore, adding the supervenience condition reveals that dynamical autonomy (as defined in the theory of differential equations) is a form of generalised autonomy, where the macrodependency is a dynamical dependency. x is unconditionally relevant for $\frac{dy}{dt}$ but, conditional on the differential equation (and so current value of y), x

[13]Dynamical systems theory allows a generalisation from the time parameter being continuous to discrete, but discussion of this is not within the scope of my current argument.

[14]Of course, this definition of dynamical autonomy is limited to descriptions that involve differential equations. One can imagine extending this in a metaphorical way, since there is a sense in which 'change over time' of features is the concern of scientific theories other than those that use differential equations. However, by extending in this metaphorical way, we lose the crispness of the definition.

is irrelevant.

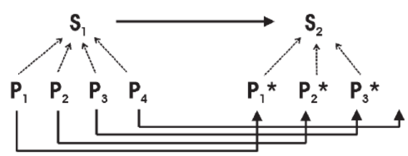

Figure 4: Fodor (1968)'s diagram. To translate to our earlier terminology: each of P_1, $P_2 ... P_4$ represent one microstate $b_1, b_2 ... b_N$ and $P{*}_1$ is time-evolved state from P_1 (in the earlier notation: $T(b)$). S_1 represents B and S_2 represents A. Earlier the arrows were uninterpreted but here we see that the vertical arrows represent supervenience relations, and the horizontal arrows represent dynamical evolution.

This dynamical autonomy (as a form of generalised autonomy) is exactly what is expressed by Fodor's famous diagram, shown in Figure 4. In this case, if we have a differential equation for S of the form $\frac{\partial S}{\partial t} = \gamma S$ (rather than $\frac{\partial S}{\partial t} = \gamma S + f(P)$) then S is dynamically autonomous of P, even though S supervenes on P, since $\frac{\partial S}{\partial t} = f(S)$.

Other names for this set up abound in the literature: Butterfield (2012) calls the commuting situation of Fodor's diagram 'meshing dynamics', List (2019) calls it 'levels of dynamics'. The key idea is that the macrovariables depend on the microvariables, but that there is nonetheless a dynamical rule describing the evolution of the macrovariables solely in terms of those variables, i.e. without having to refer back down to the microdynamics. One classic case is coin tossing. Whether the later macrostate of the coin is 'heads' or 'tails' of course depends on the fine-grained microstate, which specifies the values of the position and momentum microvariables. But there's also a macrodependency: the coin has a 50% chance of landing heads. This example demonstrates how the macrodynamics can be qualitatively different from the microdynamics, in the coin-tossing case: they are probabilistic rather than deterministic (List, 2019). In the following case study, the macrodynamics differ from the microdynamics in another respect; the macrodynamics are time-asymmetric whereas the microdynamics are time-symmetric.

6.1 An example of dynamical autonomy

In the definition of generalised autonomy, the microfacts formed an equivalence class. But this was just a formal device, and the members of the equivalence class need not be especially unified, nor belong to the same theory.

In the case of dynamical autonomy, a fuller characterisation of these equivalence classes is available. The equivalence classes may be cells of a partition over an available state space. For instance, the microstates of a gas with similar momenta form an equivalence class to give rise to a similar macrostate, e.g. temperature.

There is a framework in statistical mechanics which allows the time-irreversible equations of statistical mechanics to be constructed from the underlying time-reversible microdynamics (Wallace, 2011; Zeh, 2007; Zwanzig, 1961). The reversible microdynamics describes the time evolution of the full probability distribution ρ, $\frac{\partial\rho}{\partial t}$. In contrast, the irreversible macrodynamics describe the evolution of a different — albeit related — probability distribution, a coarse-grained distribution ρ_{cg}. These two probability distributions are related by a coarse-graining projector $\hat{P}$, which throws away information from ρ to give the 'coarse-grained probability distribution' ρ_{cg}. As such, a coarse-graining is a partition over the microstates, and so in this case the equivalence classes of microfacts have a more physical interpretation than in the general case. For example, the coarse-graining might throw away information about certain correlations between particles, and so members of a given class can differ with respect to these correlations but not with respect to, e.g. the one-particle marginal. One feature of the relationship between ρ and ρ_{cg} is particularly clear: ρ_{cg} supervenes on ρ.

In general, we might discover the macrodynamics independently from discovering the microdynamics, but in the case of this framework, we can construct the macrodynamics from the microdynamics. Just rearranging the microdynamics for the variable ρ_{cg} gives an equation of the form $\frac{\partial\rho_{cg}}{\partial t} = f(\rho_{cg}, \rho, t)$, which is not a case of autonomous dynamics.[15] But if two assumptions hold (a Markovian approximation and an initial condition cf. Wallace (2011); Robertson (2020) for more details), we can find an autonomous differential equation for the evolution of ρ_{cg} which has the form $\frac{\partial\rho_{cg}}{\partial t} = f(\rho_{cg})$.

These two assumptions are crucial to finding the autonomous dynamics. When do they hold? Are they easily fulfilled? Finding autonomous dynamics is hard; not *every* variable will be describable using an autonomous equation, because sometimes the information thrown away by coarse-graining might be dynamically relevant. Uffink (2010, p.195) says: "it is 'the art of the physicist' to find the right choice" of higher-level variable. In this sense, autonomous dynamics are rare.

In the case at hand, we have a type of meshing dynamics. Since Figure 5 commutes, we have two routes to the later coarse-grained state $\rho_{cg}(t_1)$ (macrostate A in our earlier terminology). Either we use the macrodynamics C to evolve $\rho_{cg}(t_0)$ (macrostate B) to $\rho_{cg}(t_1)$ (macrostate A); or, we use the microdynamics U to evolve $\rho(t_0)$ (microstate b) to $\rho(t_1)$, and then apply the coarse-graining, $\hat{P}$. Whilst the full description ρ and the microdynamics are unconditionally relevant, the macrodynamics screen these microdetails off from the coarse-grained state $\rho_{cg}(t_1)$ (macrostate A). We

[15] Here is the full equation:

$$\frac{\partial\rho_{cg}(t)}{\partial t} = \hat{F}\rho_{ir}(t_0) + \int_{t_0}^{t} dt'\hat{G}(t')\rho_{cg}(t-t') \tag{1}$$

where $\rho_{ir} := \rho - \rho_{cg}$, i.e. if the coarse-grained distribution is the 'relevant' distribution then ρ_{ir} is the irrelevant distribution. Here ρ_{cg} is not dynamically autonomous of the full distribution ρ, in particular it is not autonomous of ρ_{ir}, as seen in the second term. An equation of the form $\frac{d\rho_{cg}}{dt} = f(\rho_{cg}, \rho_{ir}, t)$ is not autonomous, as per the definition above.

$$\begin{tikzcd}
\rho_{cg}(t_0) \arrow[r, "C"] & \rho_{cg}(t) \\
\rho(t_0) \arrow[r, "U"] \arrow[u, "\hat{P}"'] & \rho(t) \arrow[u, "\hat{P}"']
\end{tikzcd}$$

Figure 5: A version of Fodor's diagram, where U represents the reversible microdynamics, C represents the irreversible autonomous macrodynamics and $\hat{P}$ represents coarse-graining (a many-to-one map).

can explicitly see that this is a case of generalised autonomy:

- The full-probability distribution $\rho(t_0)$ is unconditionally relevant for specifying relevant probability distribution at t_1: $P(\rho_{cg}(t_1)|\rho(t_0)) > P(\rho_{cg}(t_1))$.
- But when we have an autonomous equation for ρ_{cg}, the further details of ρ are conditionally irrelevant: $P(\rho_{cg}(t_1)|\rho_{cg}(t_0)) = P(\rho_{cg}(t_1)|\rho_{cg}(t_0)\&\rho(t_0))$.

This example demonstrates one sense in which autonomy is partial. The macrodynamics of statistical mechanics (differential equations such as the Fokker-Planck equation, Boltzmann equation, etc.) are dynamically autonomous of *some* lower-level details. But it is clear that some lower-level details are really crucial: in statistical mechanics the initial conditions are central to having the 'meshing dynamics' situation. If there are finely-balanced correlations in the initial state then the macroevolution does not dynamically decouple; the lower-level details remain important, and there is no generalised autonomy. In other words, the lower level needs to be suitably well-behaved.

In the current example, we know the condition required for the lower level to be well-behaved. In particular, we know which background conditions we need in order to rule out the deviant microstates: the initial state must be 'simple'[16] and the macrodynamics will only be empirically successful over certain timescales. The initial state condition explicitly rules out the deviant microstates, and, because of the Markovian approximation, we know to expect the macrodynamics to be empirically successful only for times less than the recurrence time.

7 Causal autonomy

Causal autonomy is controversial, not least because it depends on one's preferred account of causation. Here I steer clear of the most treacherous territory — novel causal

[16]This 'simplicity' is relative to the specific coarse-graining, so the dynamics and the coarse-graining have to mesh together: see Wallace (2011, 2012, 2013) for a thorough discussion, and connection to Albert's Past Hypothesis.

powers and mental causation — and focus only on interventionist (i.e. structural equations modelling) accounts of causation, according to which 'making a difference' is key to causation (Woodward, 2005; Pearl, 2009).

As a reminder: in the interventionist framework, one constructs a causal model $\mathcal{M}(\mathcal{S}, \mathcal{F})$, which specifies the set $\mathcal{S}$ of variables X, Y; the values those variables can take, $X : \{x_1, x_2, ...x_n\}$, $Y : \{y_1, y_2, ...y_m\}$; and the relationships between these variables as captured by structural equations $\mathcal{F}$. Then, in the interventionist framework: "X causes Y in N if and only if there are distinct values of X, x_1 and x_2, with $x_1 \neq x_2$ and distinct values of Y, y_1 and y_2 with $y_1 \neq y_2$ and some intervention such that if that intervention were to change the value of X from x_1 to x_2, then Y would change from y_1 to y_2" (Woodward, 2018, p.242).

Causal modelling is used across the special sciences from epidemiology, politics and economics to biology and quantum mechanics (cf. *inter alia* Robins et al. (2000); Pugesek (2003); Bright et al. (2016); Shrapnel (2019)). There is a plethora of different possible causal models for a given system, and some of these can be thought of as being higher- or lower-level models.[17] For instance, one model might use macrovariables whilst another uses microvariables.

A popular toy example is a pigeon pecking when represented with a red target. Call this model 1: The red colour of the target causes the pigeon to peck, and there are two variables RED:{red, not-red} and the effect variable PECK: {peck, not-peck}.

But there is another, related causal model 2 we could construct by replacing the variable RED with another variable SCARLET:{scarlet, not-scarlet}. But whilst the variable is in a sense more detailed, this causal model 2 is worse than 1, in that it fails to capture some dependencies: namely that when the pigeon is presented with a red but non-scarlet target, it will peck. This situation can be rectified by moving to a new causal model 2*, where the variable SCARLET is changed to the 'microvariable' G: {scarlet, maroon....cyan, aquamarine} and includes a functional relationship between pecking and G (Woodward, 2018, p.248).

Here we have two distinct models, or two levels of description, and we can see that generalised autonomy applies here (unsurprisingly, since it is a generalisation of Woodward's account of explanatory autonomy), as follows.

The variable G is unconditionally relevant for pecking. That is, the value of the variable G matters to whether the pigeon pecks. But conditional on the value of the variable RED, the value of G is irrelevant. Provided that the target is red, the shade of red does not matter for the pigeon.

- Unconditional relevance: $P(\text{peck} \,|\, \text{G}) > P(\text{peck})$ but
- Conditional irrelevance: $P(\text{peck}|\text{red and G}) = P(\text{peck} \,|\, \text{red})$.

This is the screening off condition: the value of RED makes the value of G irrelevant. The macrodependency between 'redness' and pecking is a causal one, and so this case of generalised autonomy is also one of causal autonomy.

[17] As such, this approach takes for granted that there exist higher-level causes, i.e. the causes don't all live at the fundamental level.

Here we can revisit the example from thermodynamics. If the macrovariable Y_t is the temperature of the water, then, conditional on the value of this variable, the microvariables (velocities of all the molecules, X_b) are irrelevant for considering the boiling point of water. In this sense, the boiling point of water is causally autonomous of the molecular motion conditional on the temperature.

We can find examples across the special sciences:

- Statham (2017, p.4826) explicitly models chemical reactions such Fisher esterifications using the structural equations framework: by removing (water) or adding (an excess of alcohol) as a starting material the reaction is driven towards the formation of an ester. The Schrödinger equation is of course relevant to the formation of chemical bonds, but the descriptions of reactions in organic chemistry need not discuss this.
- In economics, the law of demand (that a price increase will diminish the quantity demanded) is not just stating a functional relationship but the causal consequences of price changes. The macrodependency, or causal law, supervenes on the actions of individuals and their preferences, beliefs, desires and neurophysiology and so whilst the latter are unconditionally relevant, they are conditionally irrelevant.
- In biology, quantum mechanics is unconditionally relevant for a causal model of photosynthesis (since it is required for the light harvesting mechanism amongst other things), but for modelling the production of sugar it is irrelevant.

In the pigeon example, the screening off is exact. But the conditional irrelevance — as discussed in section 4.2 — is sometimes only approximate (though, as we saw, there are various strategies for restoring exactness).

7.1 Approximate screening off, and multi-level causal modelling

Imagine that Sophie the pigeon pecks red targets, unless the light reaching her eye is one particular wavelength, 653nm, in which case she does nothing. This wavelength is one of various microstates that subvene the red colour of the target, and is an example of a deviant microstate that renders 'conditional irrelevance' approximate. The previously described strategy of employing background conditions to rule out deviant microstates applies here, but an interesting issue specific to causal modelling also arises.

In the causal modelling framework, B is a cause of A (sometimes qualified with a 'relative to a model' condition cf. Statham (2018)), if there's a directed acyclic graph (satisfying the Markov and minimality condition) over a 'causally sufficient variable set' in which there is an edge between B and A. The causal sufficiency of the variable set ensures that we include the right variables (e.g. common causes) so that we don't end up with impoverished variable sets that lead us to conclude, say, that rising ice cream sales cause a rise in suncream sales, because we have neglected to include the weather as a pertinent variable.

Each of the two levels of descriptions has their own causal model (since variables related by supervenience are excluded from Woodward's account by the independent fixability condition). But the interventionist framework can be liberalised to a multi-level setting (Stern and Eva, 2020) where, in addition to causal parents, variables can also have supervenience parents. Stern and Eva discuss possible alternatives to the causal sufficiency principle as a method of determining an appropriate variable set in this context. If the microvariables (the supervenience parents) must always be added to a multi-level model, the claim that RED is the cause of PECK will be outcompeted by G (since there will no longer be an edge between RED and PECK, but instead only between G and PECK), and we run head first into the causal exclusion problem.

To avoid this, Stern and Eva (2020) suggest an approach using a condition called 'difference-maker sufficiency', according to which a variable should be included in the set if it *makes a difference*. In the case where G is exactly screened off by RED (i.e. the conditional irrelevance condition is exact, rather than approximate), G makes no difference and so needn't be added to the variable set.

In contrast, consider an example where the supervenience parents must be added to the variable set (Stern and Eva, 2020, p.17). If we start with two variables TC: total cholesterol and HD: heart disease, it looks as if TC is a cause of HD, since heart disease is positively correlated with total cholesterol (Spirtes, 2004). Total cholesterol supervenes on low-density plus high-density cholesterol. But the proportion of each type of cholesterol an individual has is important, since high-density cholesterol inhibits (and so is negatively correlated with) heart disease, whereas low density cholesterol promotes (is positively correlated with) heart disease. Since the value of which cholesterol type is present makes a difference to the effect under consideration (heart disease), the supervenience parents must be added.

The example of the pigeon pecking *unless the wavelength is 653 nm* raises the following worry. The deviant microstate (653nm) does make a difference — albeit a small one. Does this mean that the microvariable (wavelength) must be added to the variable set? And more generally, does this mean that every time the screening off is not exact, we must add in the microvariables?

If the answer is yes, this is bad news. Adding in the supervenience parents will rob the macrovariables of their causal power. We would lose the claim that B is cause of A, since the edge between them would be removed by the addition of the supervenience parents (i.e. the microvariables).[18] This would lead to the conclusion that there are no causal dependencies in the special sciences. Not only do we encounter the causal exclusion problem, this also makes the explanatory puzzle more intractable. If the higher-level causes (and laws and dynamics) of the special sciences are not genuine causes (laws or dynamics) but mere mirages, then the independent success and progress of the special sciences seems even more surprising.

[18]Note that, due to the minimality condition, also known as causal faithfulness, we must only include the minimum number of edges, which is why is the edge between B and A is removed in a DAG including b.

This returns us to the crux of Stern and Eva's project: considering the causal exclusion problem in an interventionist setting. Of course, an account of autonomy should not be expected to solve the causal exclusion problem, and in considering whether higher-level causes used in scientific practice are autonomous of their lower-level bases, we are presuming that there higher-level causes in the first place, which carries the assumption that some solution to the causal exclusion problem has been found. Nonetheless, happily, Stern and Eva's solution — difference-making sufficiency — lines up nicely with my overall approach to deviant microstates in my account of generalised autonomy.

The question facing the use of 'difference-making sufficiency' is: if a variable makes *any* minute difference, must it be included in the causal model's variable set?

In the case of empirically successful macrodependencies (like water boiling at $100°C$, or entropy increasing), the deviant microstates make no difference, at least to a certain degree of approximation. Furthermore, to improve the approximation we add in the requisite background conditions; consequently, either the deviant microstates are ruled out completely (and so make no difference) or model's domain is limited, for instance to certain timescales, which limits the consequences of the deviant microstates. Having the ability to improve an approximation allows one to both control and tame the approximation. In particular, we can see how much difference an approximation makes, and ultimately show that it is not crucial — because it could be improved or removed. Of course, the degree of approximation that is appropriate will depend on your purposes, as discussed earlier.

This is the key difference between the deviant microstates example and the cholesterol example. The former involves an approximation that can be improved. But there are no background conditions, caveats, or context dependencies you can add to the statement 'total cholesterol causes heart disease' which would mean that the microvariables make less difference. We can't improve the statement because it is not an approximation in the first place. The deviant microstates case is not like the cholesterol example. As such, the presence of the deviant microstates does not always mandate adding in the microvariables to the variable set of a causal model (and thus robbing the macrovariables of their causal power). The strategy for tackling the approximation in the case of generalised autonomy carries over to the case of causal autonomy.

8 Nomic autonomy

The signature of generalised autonomy is that there are macrodependencies which screen off the underlying microfacts. In the last two sections, the macrodependencies were dynamical and causal, respectively. Now the question is: can a macrodependency in a case of generalised autonomy be a law-like, or nomic, macrodependency? I think the answer is fairly clearly yes. One example is the ideal gas law[19], $pV = nRT$. Similarly to the example of the boiling kettle, this dependency is independent of the microstate specifying every molecule's position and momenta. Other examples abound — especially once causal and dynamical autonomy are on the table. The statement 'water boils at $100°C$' looks like a law, as does the law of demand and the Boltzmann equation discussed in section 6.1. Of course, the legitimacy of non-fundamental laws is disputed. but according to one influential account of laws in physics (Maudlin, 2007), a higher-level dynamical equation qualifies as a LOTE (law of temporal evolution) and so the examples of dynamical autonomy in section 6 immediately qualify as cases of nomic autonomy. Likewise, if causal generalisations qualify as laws, then the examples of the previous section are also cases of nomic autonomy.

9 Specifying the lower-level details

There is another aspect to the importance of laws in the understanding of autonomy. So far, as discussed in section 5, generalised autonomy does not dictate the nature of the conditionally irrelevant details. But in particular case studies we can look at the equivalence class of microfacts and specify what they have in common and how they differ, and thus learn *which* details do not matter.

Coin tossing dynamics are well-understood, and the macrostate (Heads/Tails: 50/50) can be constructed from the microfacts (the microdynamics describing the evolution of the position and momenta) using Poincare's method of arbitrary functions (Poincaré, 1896). This explains which details do not matter (within reasonable bounds, *how* the coin is tossed) and under which circumstances (the initial credence function needs to be suitably 'random', see Myrvold (2012); Butterfield (2011b) for more details). Thus, in this case, we can specify *which* details don't matter.

Likewise, in the macrodynamics of statistical mechanics, such as the Boltzmann equation, the details thrown away by coarse-graining, such as two-or-more particle correlations in the full probability distribution, are irrelevant. In this case it is a feature of the microstate rather than the microdynamics that doesn't matter (because it is conditionally irrelevant). But in other cases it could be that the lower-level details that don't matter are the microdynamics, or more generally the microlaws. That is, the microfacts in the equivalence class that subvenes B could differ with respect to which dynamics or laws describe how the microstate evolves over time. For example, the microlaw could be a Hamiltonian describing how the microstate moves through phase space. In renormalisation group methods, the coarse-graining is over different Hamiltonians, i.e. different microlaws, and so the resulting description is autonomous of which exact Hamiltonian describes the system. To mark this distinction, Franklin (2020) calls this latter case 'microlaw autonomy' and the former 'microstate auton-

[19]This law relates the pressure 'p', the volume 'V', the number of moles 'n', the gas constant 'R' and the temperature 'T'.

omy'. Both microlaw and microstate autonomy are subtypes of generalised autonomy that differ with respect to *which* lower-level details are conditionally irrelevant. In the microlaw autonomy case, b_1 and b_2 differ with respect to the microlaw, i.e. the Hamiltonian, and in the microstate autonomy case b_1 and b_2 differ with respect to the microstate, i.e. the exact location in phase space.

The idea that the microfacts in the equivalence class belong to distinct microtheories can be called 'microtheory autonomy'. The ideal gas law once again provides an illustration; it can be derived from either classical or quantum statistical mechanics. This is a case of generalised autonomy, which can also be categorised as microtheory autonomy, since the different microfacts in the equivalence class come from different microtheories: $P(p|\frac{nRT}{V}\&QSM \text{ description}) = P(p|\frac{nRT}{V}\&CSM \text{ description})$.[20] Here, clearly, the probabilities are to be understood as epistemic probabilities over theory space that are nonetheless objective in the manner discussed earlier.

Unsurprisingly, since it is a sub-type of generalised autonomy, microtheory autonomy is a partial notion of autonomy. For example, the ideal gas law is not fully autonomous of the lower-level theory, since some underlying theories won't give rise to the ideal gas law. So there are some lower-level descriptions or microtheories ruled out by the macrofact.

The partial nature of microtheory autonomy is crucial in accounting for the epistemic role of higher-level theories. Namely, in certain cases a higher-level theory is used to guide the search for a lower-level, i.e. more fundamental, theory. A prominent example is black hole thermodynamics, which is routinely taken to guide the search for a theory of quantum gravity. Yet if thermodynamics were fully autonomous and so 'floated free' of any underlying theory, it is hard to see how it could play this epistemic role, since thermodynamics would be compatible with *any* 'microtheory'. But since thermodynamics is only partially autonomous, it usefully limits the number of viable fundamental theories. An instance of this is the theoretical discovery that a form of string theory (M-theory) can reproduce the entropy ascribed to a black hole by black hole thermodynamics (Strominger and Vafa, 1996) being received as favourable evidence for M-theory. Not only is this one of the most highly cited articles in high energy physics, it also won the Physics Frontiers Breakthrough medal.

Other examples of higher-level theories guiding the search for more fundamental theories scatter the history of science — the ultraviolet catastrophe guiding the search for quantum mechanics is another prominent example — yet the case of black hole thermodynamics is especially striking given the frequency with which thermodynamics is proclaimed to be an autonomous theory. This only further emphasises that the notion of autonomy most useful to science is a partial one.

[20] In the ideal gas case, whilst this is autonomous from the underlying theory, it is not totally clear whether this is microlaw or microstate autonomy; the dynamics are not particularly important, since what it really comes down to is the combinatorics of states, i.e. finding the partition function which in some sense counts the number of possible energy states.

10 The explanatory power of generalised autonomy

In section 3, I outlined an explanatory puzzle: why does physics — or, more generally, why do lower-level theories — not matter more for the practice of the special sciences? The answer is that some macrofacts are autonomous in the sense of generalised autonomy. If a macrofact has generalised autonomy, then this explains why the lower-level details don't matter: they are genuinely (though conditionally) irrelevant.

Indeed, if a macrofact has generalised autonomy, then it is unsurprising that the investigation or discovery of this fact can proceed independently of the lower level. If the Schrödinger equation applies to the wildebeest migration, then it is unconditionally relevant for the wildebeest, but conditional on the evolutionary history, biological variables and the pattern of rainfall, the details of the Schrödinger equation are irrelevant. Generalised autonomy licences the claim that the details of the Schrödinger equation are genuinely irrelevant for the wildebeest migration, and thus explains why ecologists need not study quantum physics in order to make progress in predicting the pattern of movement of the wildebeest. Moreover, since generalised autonomy is a worldly feature, the methodological autonomy of ecology is not a contingent consequence of our cognitive and calculational limitations, but is instead explained by a feature of the world: generalised autonomy.

Here is another example: why could the pioneers of thermodynamics proceed so successfully despite their ignorance about the nature of matter and, in the early stages of the theory, the nature of heat[21]? The quantum nature of matter is unconditionally relevant to the ideal gas law, but is irrelevant conditional on other thermodynamic variables like temperature and volume. As such, the ideal gas law is nomically autonomous. We saw that a further subtype of autonomy applies to the gas law: microtheory autonomy. The ideal gas law is autonomous of whether the underlying theory is quantum statistical mechanics or classical statistical mechanics; this explains why ignorance of the quantum nature of matter was no impediment to the pioneers of thermodynamics. Of course, generalised autonomy is only a partial notion of autonomy. Some microtheories wouldn't give rise to the ideal gas law. Furthermore, there may be some macrofacts related to the ideal gas law which do not have generalised autonomy; the description of cold, dense gases requires quantum mechanics, for example.

Generalised autonomy is a partial autonomy. Does this place limits on the ability of generalised autonomy to explain the methodological independence of the special sciences? I think not, because where there is no generalised autonomy, there is no methodological independence. Earlier we saw that whilst the macrofacts about the migration of the European robin have generalised autonomy, macrofacts about the mechanism for navigation do not, since they require quantum mechanics. This is also a case where biology is *not* methodologically autonomous of physics (hence the name:

[21]Carnot, when developing his eponymous Carnot cycle, believed that heat was a fluid called caloric, and modelled the cycle on water. Upon the discovery that heat is 'nought but molecules in motion' (Uffink, 1996) surprisingly little needed to be altered about the Carnot cycle (Cercignani, 1998).

quantum biology), so the lack of generalised autonomy explains the lack of methodological autonomy. Generalised autonomy explains why methodological autonomy is such a successful strategy and consequently, in cases where generalised autonomy does not hold, we should not immediately expect to see successful methodological autonomy.

11 Conclusion

In this paper I have offered an account of autonomy called generalised autonomy, according to which a macrofact A has generalised autonomy when a microfact b is unconditionally relevant for A but is screened off by another macrofact B, thus making this underlying microfact conditionally irrelevant. This account treads a third way between Fodor's weak methodological autonomy and his implausibly strong metaphysical or full autonomy, in which God is free to choose the macrofacts at will. I argued that 'full autonomy' is too strong to be of any use to understanding the special sciences, and so we should retire the God metaphor. Autonomy is not all or nothing, but partial, in the generalised autonomy account.

Generalised autonomy unifies dynamical, causal and nomic autonomy, according to the type of macrodependency a given example exhibits. But determining *which* lower-level details are conditionally irrelevant will be a case by case matter and needn't be specified for any given case of generalised autonomy. Consequently, generalised autonomy is independent of whether inter-theoretic reduction holds.

However, there is one further distinction useful for special science theories in physics: sometimes it doesn't matter which microtheory underlies the macrofact. I have called this microtheory autonomy. As a subspecies of generalised autonomy, this is a partial autonomy; not *any* microtheory is compatible with the macrofacts, even if these macrofacts have generalised autonomy. This partial nature of autonomy is central for accounting for the epistemic role played by macrofacts in searching for more fundamental theories. Finally, generalised autonomy explains why the microfacts described by fundamental physics (or a lower-level theory) do not matter for practitioners of special sciences — despite being unconditionally relevant. Once you have your hands on the macrodependency, the microdetails are conditionally irrelevant.

References

Anderson (1972). More is different. *Science*, 177(4047).

Batterman, R. (2018). Autonomy of theories: An explanatory problem. *Noûs*, 52(4):858–873.

Bright, L. K., Malinsky, D., and Thompson, M. (2016). Causally interpreting intersectionality theory. *Philosophy of Science*, 83(1):pp. 60–81.

Butterfield, J. (2011a). Emergence, reduction and supervenience: a varied landscape. *Foundations of Physics*, 41(6):920–959.

Butterfield, J. (2011b). Less is different: emergence and reduction reconciled. *Foundations of Physics*, 41(6):1065–1135.

Butterfield, J. (2012). Laws, causation and dynamics at different levels. *Interface Focus*, 2(1):101–114.

Cartwright, N. (1983). *How the Laws of Physics Lie*. Oxford University Press, Oxford.

Cartwright, N. (1999). *The dappled world: A study of the boundaries of science*. Cambridge University Press, Cambridge.

Cercignani, C. (1998). *Ludwig Boltzmann: the man who trusted atoms*. Oxford University Press, Oxford.

Earman, J. and Roberts, J. T. (1999). "Ceteris Paribus", there is no problem of provisos. *Synthese*, 118(3):439–478.

Fodor, J. A. (1968). *Psychological explanation: An introduction to the philosophy of psychology*. Random House, New York.

Fodor, J. A. (1974). Special sciences (or: The disunity of science as a working hypothesis). *Synthese*, 28(2):97–115.

Franklin, A. (2020). Whence the effectiveness of effective field theories? *British Journal for the Philosophy of Science*, 71(4):1235–1259.

Handfield, T. and Wilson, A. (2014). Chance and context. In *Chance and Temporal Asymmetry*. Oxford University Press, Oxford.

Heil, J. (1998). Supervenience deconstructed. *European Journal of Philosophy*, 6(2):146–155.

Hellman, G. P. and Thompson, F. W. (1976). Physicalism: Ontology, determination, and reduction. *The Journal of Philosophy*, 72(17):551–564.

Hemmo, M. and Shenker, O. (2012). *The Road to Maxwell's Demon: Conceptual Foundations of Statistical Mechanics*. Cambridge University Press, Cambridge.

Hoefer, C. (2007). The third way on objective probability: A sceptic's guide to objective chance. *Mind*, (116):549–96.

Kim, J. (2000). *Mind in a physical world: An essay on the mind-body problem and mental causation*. MIT press, Cambridge, Mass.

Lewis, D. (1986). A subjectivist's guide to objective chance. In *Philosophical Papers vol. II*, pages 83–133. OUP, Oxford.

List, C. (2019). Levels: descriptive, explanatory, and ontological. *Noûs*, 53(4):852–883.

List, C. and Pivato, M. (2015). Emergent chance. *Philosophical Review*, 124(1):119–152.

Loewer, B. (2008). Why there is anything other than physics. In Hohwy, J. and Kallestrup, J., editors, *Being Reduced: New Essays on Reduction, Explanation, and Causation,*. Oxford university press, Oxford.

Loewer, B. (2009). Why is there anything except physics? *Synthese*, 170(2):217–233.

Martin, J. (2015). Fundamental disputations: The philosophical debates that governed american physics, 1939–1993. *Hist Stud Nat Sci*, 45(5):703–757.

Maudlin, T. (2007). *The Metaphysics within Physics*. Oxford University Press.

McLaughlin, B. (1992). The rise and fall of british emergentism. In Ansgar Beckermann, H. F. . J. K., editor, *Emergence or reduction? The prospects for non-reductive physicalism*, pages 49–93. De Gruyter, Berlin.

Myrvold, W. C. (2012). Deterministic laws and epistemic chances. In Yemima Ben-Menahem, M. H., editor, *Probability in Physics*, pages 73–85. Springer, New York.

Nagel, E. (1961). *The Structure of Science: problems in the Logic of Scientific Explanation*. Harcourt, San Diego, California.

Oppenheim, P. and Putnam, H. (1958). The unity of science as a working hypothesis. In Feigl, H., editor, *Minnesota Studies in the Philosophy of Science*, volume 2. Minnesota University Press, Minneapolis.

Pearl, J. (2009). *Causality: Models, Reasoning, and Inference*. Cambridge University Press, Cambridge, second edition.

Poincaré, H. (1896). *Calcul de Probabilities*. Gauthier-Villars, Paris.

Potochnik, A. (2017). *Idealization and the aims of science*. University of Chicago Press, Chicago.

Pugesek, B., T. A. . V. E. A., editor (2003). *Structural Equation Modeling: Applications in Ecological and Evolutionary Biology*. Cambridge University Press, Cambridge.

Reutlinger, A., Schurz, G., Hüttemann, A., and Jaag, S. (2020). Ceteris Paribus Laws. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University.

Robertson, K. (2020). Asymmetry, abstraction and autonomy: how to justify coarse-graining in statistical mechanics. *British Journal for the Philosophy of Science*, 71(2):547–579.

Robins, J., Hernán, M., and Brumback, B. (2000). Marginal structural models and causal inference in epidemiology. *Epidemiology*, 11(5):550–60.

Robinson, J. C. (2004). *An introduction to ordinary differential equations*. Cambridge University Press, Cambridge.

Rodgers, C. T., . H. P. J. (2009). Chemical magneto reception in birds: the radical pair mechanism. *Proceedings of the National Academy of Sciences*, 106(2):353–360.

Schaffer, J. (2007). Deterministic chance? *British Journal for the Philosophy of Science*, 58:113–140.

Schaffer, J. (2016). Grounding in the image of causation. *Philosophical Studies*, 173(1):49–100.

Schrödinger, E. (1967). *What is life?: the physical aspect of the living cell, and, Mind and matter.* Cambridge University Press, Cambridge.

Shrapnel, S. (2014). Quantum causal explanation: or, why birds fly south. *European Journal of Philosophy of Science*, 4:409–423.

Shrapnel, S. (2019). Discovering quantum causal models. *British Journal for the Philosophy of Science*, 70(1):1–25.

Sklar, L. (1993). *Physics and chance: Philosophical issues in the foundations of statistical mechanics*. Cambridge University Press, Cambridge.

Sklar, L. (2015). The philosophy of statistical mechanics. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University.

Sober, E. (1999). The multiple realizability argument against reductionism. *Philosophy of science*, 66(4):542–564.

Spirtes, P., . S. R. (2004). Causal inference of ambiguous manipulations. *Philosophy of Science*, 71(5):833–845.

Statham, G. (2017). The manipulation of chemical reactions: probing the limits of interventions. *Synthese*, 194:4815–4838.

Statham, G. (2018). Woodward and variable relativity. *Philosophical Studies*, 175:885–902.

Stern, R. and Eva, B. (2020). Antireductionist interventionism. *British Journal for the Philosophy of Science*.

Strominger, A. and Vafa, G. (1996). Microscopic origin of the Bekenstein-Hawking entropy. *Physics Letters B*, 379:99–104.

Uffink, J. (1996). Nought but molecules in motion. *Studies in History and Philosophy of Modern Physics*, 27(3):373–387.

Uffink, J. (2010). Irreversibility in stochastic dynamics. In Gerhard Ernst, A. H., editor, *Time, chance and reduction: philosophical aspects of statistical mechanics*, pages 180–207. Cambridge University Press, Cambridge.

Wallace, D. (2011). The logic of the past hypothesis. http://philsci-archive.pitt.edu/8894/.

Wallace, D. (2012). The arrow of time in physics. In Bardon, A. and Dyke, H., editors, *A Companion to the Philosophy of Time*. Wiley-Blackwell, Chichester, West Sussex; Malden, MA.

Wallace, D. (2013). Inferential vs. dynamical conceptions of physics. *arXiv preprint arXiv:1306.4907*.

Wallace, D. (2019). Naturalness and emergence. *The Monist*, 102(4):499–524.

Woodward, J. (2005.). *Making things happen: A theory of causal explanation.* Oxford university press, Oxford.

Woodward, J. (2018). Explanatory autonomy: the role of proportionality, stability, and conditional irrelevance. *Synthese*, pages 1–29.

Zeh, H. D. (2007). *The physical basis of the direction of time*. Springer, Berlin, 5th edition.

Zwanzig, R. (1961). Statistical mechanics of irreversibility. In *Lectures in Theoretical Physics (Boulder), Vol. 3*, pages 106–141. Inter-science, New York.