Isolated Systems and their Symmetries, Part I: General Framework and Particle-Mechanics Examples

David Wallace

October 21, 2021

Abstract

Physical theories, for the most part, should be understood as modelling isolated subsystems of a larger universe; doing so, among other benefits, greatly clarifies the interpretation of the dynamical symmetries of those theories. I provide a detailed framework for analysing the subsystem structure of physical theories and applying it to the interpretation of their symmetries: the core concept is *subsystem recursivity*, whereby interpretative conclusions about a sector of a theory can be deduced from considering subsystems of other models of the same theory. I illustrate the framework by extensive examples from nonrelativistic particle mechanics, and in particular from Newtonian theories of gravity. A sequel to the paper will apply the framework to the local and global symmetries of classical field theory.

1 Introduction: Symmetry and the Cosmological Assumption

> [I]t is not often that experiments are done under the stars. Rather they are done in a room. Although it is physically reasonable that the walls have no effect, it is true that the original problem is set up as an idealization.
>
> Richard Feynman.[1]

Take a dynamical symmetry of a theory — translational symmetry in Newtonian mechanics, say, or phase symmetry in the complex Klein-Gordon theory. Here are four theses[2] one might defend about the symmetry:

1. The Unobservability Thesis: symmetry-variant features of a theory are unobservable.

[1] Feynman and Hibbs (1965, p.94).

[2] I draw all four from Wallace (2019b), where each is discussed in more detail.

2. The Representational Equivalence Thesis: symmetry-related models of a theory are equally apt to represent a given state of a given physical system.

3. The Modal Equivalence Thesis: given a model used to represent a system, any other model obtained from the first by a symmetry transformation re-represents the same system in the same state, rather than representing a different possible state of that system.

4. The Surplus Structure Thesis: mathematical structure used in the statement of a theory, but variant under the symmetries of that theory, should not be thought of as representing anything physically significant.

All four are contentious, and all four have been extensively discussed in the recent literature. To give a sample of the issues raised:

- If the notion of symmetry under consideration is indeed dynamical — that is, a symmetry is a transformation that takes solutions of the equations of motion to other such solutions — then it is not easy to see why this no-doubt-practically-useful fact about a theory should have implications for what is observable, or what is physically real. (Hence Ismael and van Fraassen (2003), Allori *et al.* (2008), and Dasgupta (2016) all replace or supplement the dynamical definition of symmetry with a definition that makes direct reference to what is observable.)

- The notion of dynamical symmetry seems far too liberal for us to accept the Representational Equivalence Thesis without qualification. All manner of physical systems admit dynamical symmetries which pre-theoretically clearly do not relate representationally equivalent models (for instance,[3] the Lenz-Runge symmetry of the two-body problem includes symmetries which change the shape of a body's orbit; the conformal symmetries of vacuum electromagnetism can change the wavelength of an electromagnetic wave). As Ismael and van Fraassen (2003) and Belot (2013) argue at length, again we seem to need some substantial restriction in the definition of 'symmetry'.

- Philosophers have generally adopted an 'all-or-nothing' approach to modal equivalence: most philosophers who have defended some form of the doctrine (see, e. g. , Saunders 2003) have defended it for all symmetries, and the view that 'local' symmetries like the diffeomorphism symmetry uniformly obey Modal Equivalence is close to orthodoxy. But as Belot (2018) points out, this does not conform to physics usage: physicists normally regard 'global' symmetries as generating real physical differences, while local symmetries do in some circumstances and not in others, according to what their form is at spatial infinity.

The vast majority of these discussions — as with the vast majority of all discussions of theory interpretation in philosophy of physics — tacitly or explicitly make two interpretative assumptions:

[3] Both examples are discussed in detail in (Wallace 2019b).

The Fundamentality Assumption: The primary way to understand a physical theory is to suppose that it is exactly true and represents the deepest features of reality. Interpretations based on the assumption that the theory is approximate, or otherwise non-fundamental are secondary and derivative on the primary understanding, or else of merely instrumental significance.

The Cosmological Assumption: The primary way to understand a physical theory is to suppose that it models the entire universe. Interpretations based on the assumption that the theory models a finite subsystem of a larger universe are secondary and derivative on the primary understanding, or else of merely instrumental significance.

The Fundamentality Assumption is ubiquitous in philosophy of science (consider van Fraassen's (1980, p.8) influential definition of realism: "acceptance of a scientific theory involves the belief that it is true") and is widely assumed in philosophy of physics. To give a few examples (mostly borrowed from a more extensive list collected by Williams (2019)): "Whatever else it means to interpret a scientific theory, it means saying what the world would have to be like if the theory is true" (Earman 2004, p.1234); "To interpret a theory is to describe the possible worlds about which it is the literal truth" (Belot 1998, p.532); "The content of a theory is given by that set of worlds of which the theory is true". (Ruetsche 2011, p.6).

The Fundamentality Assumption leads to trouble in the interpretation of physical theories, for instance in quantum field theory (as discussed in Williams, *ibid*) and in the quantum measurement problem (Wallace 2020b). But in this paper I will be concerned more with the Cosmological Assumption, in particular in the context of theories set on a (background or dynamical) spacetime, where the assumption amounts to supposing that the theory should in the first instance be understood as a theory of the *entire* spacetime. It is somewhat less common to see it stated explicitly, but it plays a large role in, e. g. , discussions of the ontology of quantum mechanics (where talk of 'the wavefunction of the universe' is common and the wavefunction of subsystems is secondary and derivative; see in this context (Albert 1996) and the various contributions to (Ney and Albert 2013)), and in considerations of Leibniz shifts, of background independence, and of the Hole Argument in the philosophy of spacetime (for instance, see Belot (2011, p.2866) for an explicit restriction of the discussion to theories of the whole universe).

To reject the Fundamentality Assumption is to accept that (many or most) physical theories can and should be directly interpreted as modelling a system only to a certain degree of approximation and with respect to certain features; to reject the Cosmological Assumption is to accept that (many or most) physical theories can and should be directly interpreted as modelling subsystems of a larger reality. The two are not fully separate: to understand a theory as modelling a proper subsystem is to accept that the theory is not in itself a full account of reality and must be embedded in some larger theory if the relation between the subsystem and its surroundings is to be understood; conversely,

once we understand a theory as modelling a system only with respect to certain features it is a short step to recognize that those features are likely to be defined on a system-by-system basis and not for the universe as a whole.

To see what is wrong with both assumptions, let's consider the example of N-particle classical mechanics. We can't even *ask* what the world would be like according to this theory unless we know the value of N. Sometimes (mostly in discussions of the interpretation of quantum mechanics) it's suggested that we should take N to be the number of elementary particles in the observable universe (depending on how you count, $\sim 10^{70}$). But it's opaque what the motivation for that is: in the actual universe those elementary particles are mostly highly-relativistic particles with irreducibly quantum dynamics, so the fictional universe in which there are exactly that many particles but they obey Newtonian equations of motion has little or nothing to do with the actual universe.

But in any case, none of the actual uses of Newtonian particle mechanics in physics assume *either* that $N \sim 10^{70}$, *or* that the 'particles' are elementary. The classic applications of the theory took the 'particles' to be planets, moons or comets, and N to be 2 (for the Earth-Moon system, or for a body orbiting the Sun), or to be the number of planets in the Solar system; more modern applications (given access to modern computers) might take the 'particles' to be stars, or galaxies, or agglomerations thereof and so take N to be in the millions or billions. It's clear that in all these applications, (a) the 'particles' do not represent elementary particles, but represent the collective degrees of freedom of larger bodies; (b) the theory is being applied to some subsystem of the universe. Furthermore, the fact that N is a freely-choosable variable in Newtonian physics is useful in applications of the theory precisely because we want to apply it to different physical systems with different numbers of constituents.

That being the case, there is no obvious naturalistic motivation to understand Newtonian mechanics in the first instance as an exactly-true theory of an entire universe, rather than to seek an understanding of the theory where its models are to be understood as models of isolated subsystems. And from that perspective, the fact that the environments of those systems have been removed to infinity is an idealization, not something to be understood in physical terms.

In considering the epistemology and metaphysics of symmetry, whether or not we adopt the Cosmological Assumption has major implications. Notably:

1. Observations are physical processes, but they are not normally modelled explicitly within the system being studied, but are considered as external interventions. And then it becomes relatively easy to see when the presence of a dynamical symmetry has implications for observability of physical quantities: it can do so *when the symmetry can be extended so as to apply also to the dynamics of those interventions.*

2. The Modal Equivalence Thesis has a rather metaphysical — even esoteric — character when it is applied to models of the entire universe (as can be seen by the recent literature, with its discussion of counterparts, haecceitism and the like). Even Belot's (2018) naturalistic argument against

> unrestricted Equivalence is largely a matter of conventionalism and conceptual analysis: the reason we should reject unrestricted Equivalence is that most physicists do and they ought to be taken as authoritative as to what their words mean. But when applied to subsystems of the universe, it becomes relatively simple to understand these modal questions in more directly empirical terms: is a situation where the symmetry transformation is applied to this system, but not to other systems, the same as or different from the original situation?

(I should note that the idea that transformations to subsystems give rise to modally, and indeed empirically, distinct states of affairs is widely recognized and has been extensively discussed: see, e. g. , (Brown and Sypel 1995), (Healey 2009), (Greaves and Wallace 2014), and (Belot 2018). But for the most part these discussions still adhere to the Cosmological Assumption in that they take for granted the physical theories should *in the first instance* be understood as modelling the whole universe, with the application of symmetry transformations to subsystems having a derivative character. This is fairly explicit in Belot's, and Greaves and Wallace's, papers.)

A major assumption of this paper is that the interpretative questions of symmetry *cannot* meaningfully be answered as long as we adhere to the Cosmological Assumption, because their answers depend not just on how the symmetry applies to the system itself but on whether and how it can be extended to surrounding or distant systems. In a companion paper (Wallace 2019b), indeed, I explore this at length and offer arguments for (partial versions of) the Unobservability, Representational Equivalence, Surplus Structure, and Modal Equivalence theses, depending in each case on just how a symmetry extends from a system to other systems.

Here, though, I am concerned with a slightly different puzzle. The attraction of symmetry-based arguments is that they start with something formal — the notion of a transformation, perhaps of a certain formally-characterised kind, that leaves the dynamics invariant — and draw substantive, interpretative conclusions from that formal feature. This is striking (so striking, indeed, as to have aroused the suspicion that it is impossible: Dasgupta (2016) and Møller-Nielsen (2017) can both be read as voicing this suspicion). But if instead the interpretative significance of a symmetry depends on questions of how that symmetry extends beyond the system — questions which, it might seem, definitionally cannot depend on formal features of the system — then the striking feature seems mostly lost.

Remarkably, this is not the case, at least not for a great many theories of contemporary relevance. This is because these theories have the remarkable and underappreciated feature of being able to reinterpret subsystems of their models, when dynamically isolated, as other models of the same theory. Or put the other way around: a given theory's models can be interpreted, in idealization, as representing subsystems of other, more complicated, models of the same theory. I call theories with this feature *subsystem-recursive*: given a subsystem-recursive theory, we can draw conclusions about the proper subsystems of a given model

from a consideration of its dynamics, and then read those conclusions over to other models of that same system in their entirety.

My purpose here, then, is twofold: to give a more precise definition of subsystem-recursivity, and to use it to give a systematic framework for answering questions of modal equivalence and observability for theories with differing symmetry groups. (The resultant framework can be seen as a development of the earlier framework that Greaves and I developed in previous work (Greaves and Wallace 2014).) This is the first of two papers on the topic: this paper focuses on the general concept of subsystem-recursivity and on examples drawn from nonrelativistic particle mechanics. The sequel (Wallace 2019a, henceforth Part II) considers the more technically complex case of local and global symmetries in field theory, in which context the basic ideas remain valid but considerable subtleties develop.

The structure of the present paper is as follows: in section 2 I develop the idea of a 'sector' of a theory as a subclass of models of that theory intended to model different possible states of the same system (whereas a theory, understood more broadly, also can be used to model many different systems), and then introduce the key idea of subsystem-recursivity. In sections 3–4 I introduce a general framework for discussing the subsystems of a (sector of a) theory, and then develop the appropriate notion of dynamical symmetry within that framework. In the remainder of the paper sections I carry out an extended exploration of the related notions of symmetry and subsystem-recursivity in the context of non-relativistic particle theories in general (section 5) and of various forms of Newtonian gravity in particular (section 6). Section 7 is the conclusion.

As for notation: Roman indices (raised or lowered) range from 1 to 3, and normally denote components of 3-spatial vectors; repeated indices are summed over; indices can be lowered or raised at will (so that $x^i = x_i$). 3-spatial vectors themselves are denoted with arrow superscripts, as in $\vec{x}$.

2 Theories, sectors, and subsystem recursivity

The term 'theory', at least in physical practice, is ambiguous between various elements of a complex hierarchy. Take the case of N-particle mechanics: here, in ascending order of generality, 'theory' might refer to

1. N-particle mechanics with a specific value of N (say $N = 2$), a specific form of the force law between the N particles (say, a Newtonian inverse-square force, defined by a potential

$$V_{ij}(r) = -Gm_i m_j/r \tag{1}$$

where r is the interparticle distance between particles i and j, m_i and m_j are their masses, and $G = 6.67 \times 10^{-11}\text{kg}^{-1}\text{m}^3\text{s}^{-2}$), and specific choices of the parameters for the particles, in this case their masses (say, $m_1 = 1.99 \times 10^{30}$kg, $m_2 = 5.97 \times 10^{24}$kg — the masses, respectively, of Sun and Earth).

2. N-particle mechanics where the force law is held fixed, but now the number N of particles as well as the individual particles' parameters (masses, in the gravitational case) can vary between models.

3. N-particle mechanics where the parameters in the force law are allowed to vary (so that G can take on different values).

4. N-particle mechanics where even the functional form of the force law varies between models (in some models it might be a Coulombic force law where the particles are assigned charges as well as masses; in another it might include an external, frictional force law).

Similarly, if our theory is (classical) complex scalar field theory, 'the theory' might be understood as including a (1) specific choice of Lagrangian, parameters in the Lagrangian, and boundary conditions; or it might permit (2) models with different boundary conditions; or (3) models with different coefficients in the Lagrangian; or (4) models with different functional forms in the Lagrangian. (And even this is not the top of the hierarchy: if I had instead spoken of 'classical field theory', then different models could differ even in their degrees of freedom.)

All of these uses of 'theory' have value in some contexts: those contexts can usefully be understood by considering the sorts of modality that the different models represent. In case (1), different models represent different possible dynamical histories of a given system: in my particular example: the Earth-Sun system. Given a model that represents the actual way the Earth-Sun system is, other models represent the way it would be given different initial conditions (but without moving outside the space of dynamics described by the theory itself by, for instance, ablating matter off the Earth). In case (2), for the most part we are not talking about physical modality: different models represent different, related systems, themselves actual or possible. If we change m_2 to 7.35×10^{22} kg, now we have a class of models suitable for representing the Earth-Moon system; if we change N to 9 and adjust the masses appropriately, we have a class of models suitable for representing all the planets of the Solar system. Note that different theories in class (2) can be, and are, used to represent different systems in the same possible world (in particular, they can be used to represent distinct, coexisting, actual systems).

In case (3), different models for the most part represent different possible ways the whole world might be: when we imagine changing G to some larger number we are discussing a counterfactual for the whole universe, not describing a system which for all we know obtains elsewhere in the Milky Way.[4] In case (4) the same is largely true but we will also have models suitable for describing completely different systems with completely different interpretations of the model (as when we change from Newtonian gravity to Coulombic electrostatics, for instance.)

For my purposes, the relevant levels of generality are (2) and, derivatively, (1). A 'theory', in the sense relevant for my purposes, has many classes of

[4]The situation is complicated by the fact that these parameters are dimensionful, so that changes in G could also be interpreted as changes of units.

models which are candidates to represent many different systems, but all of those systems describe the same sorts of entities evolving under the same dynamics, many potentially co-existing. So in Newtonian gravity, a theory is specified by the force law and the coefficients in that law; in scalar field theory, by a Lagrangian and its coefficients. Borrowing one precisification of the somewhat-inconsistent conventions of physics, I'll say that a *sector* of a theory is a (maximal) sub-class of models all of which are candidates to represent different histories of the same system. In point-particle Newtonian gravity with the inverse-square law, a sector is specified by a choice of particle number and an assignment of a mass to each particle; in scalar field theory, a sector is specified by a choice of boundary conditions; in Newtonian gravity using a gravitational potential, specifying a sector requires both boundary conditions for the field and a choice of particle number and particle masses. Borrowing from the field case, I'll say that 'boundary conditions' are by definition what individuates sectors: this has its normal meaning in pure field theory, though it can sound a little odd in particle mechanics to regard a mass as part of a boundary condition (it might seem more natural if one thinks of the 'condition' as being imposed at some initial time, when whatever physical process that brought the system into being — planetary coalescence, say — determined those parameters). Different models in a sector then represent different possible histories for the system, individuated by, say, initial choices of position and velocity for the particles.

In both particle and field theories, there is a natural way to speak of a sector's *subsystems*: a subsystem of an N-sector of particle mechanics is a collection $M \leq N$ of the particles, perhaps over some restricted time; a subsystem of a sector of field theory is the restriction of the fields to some well-behaved subset of spacetime. (I will be more precise about these definitions in due course.) It will not in general be possible to give a self-contained dynamics for a subsystem, for the obvious reason that other parts of the larger system may influence the subsystem. But for a given *model*, it might be that some subsystems *are* dynamically isolated, in the sense that we can give an autonomous dynamics for the subsystem, with the background conditions as set by that model and independent *to some degree* from distant variations in the rest of the system.[5]

The notion is easier to illustrate through examples than to define precisely. As a concrete illustration, consider Coulombic electrostatics.[6] A sector of this theory is given by N masses $m_1, \ldots m_N$ (all positive real numbers), by N charges $q_1, \ldots q_N$, and by a boundary condition $\Phi(\infty, t)$ (a smooth real function from $\mathbb{R}$ to $\mathbb{R}$). Within each sector, the kinematic models are ordered N-tuples of

[5] I slightly abuse terminology here by using 'subsystem' both for the formal (model-based) notion of subsystem, and the physical notion, so that given the Solar system, 'subsystem' might denote either the Earth-Moon system or the 2-particle sub-model of a mathematical model of the Solar system. Thanks to Valeriya Chasova for pointing out this ambiguity; hopefully it is unambiguous in context.

[6] Why not gravitation? The equivalence principle makes the notion of 'isolated subsystem' distinctly subtle in a gravitational theory, so I postpone further discussion of this case until section 6.

trajectories $x_1^i, \ldots x_N^i$; the dynamical models in addition satisfy

$$m_J \frac{\mathrm{d}}{\mathrm{d}t} x_J^i(t) = q_J \nabla_i \Phi(x_J^i(t), t) \tag{2}$$

where $\Phi(x,t)$ is the unique solution to the equation

$$\nabla^2 \Phi(x,t) = 4\pi \sum_J \delta(x - x_J) q_J \tag{3}$$

subject to the boundary condition $\lim_{|x|\to\infty} \Phi(x,t) = \Phi(\infty, t)$.

For the moment, we can define a subsystem by a subset $M \leq N$ of the trajectories (in section 3 it will be convenient to adjust this definition slightly, to allow for subsystems of finite temporal extent) but the definition does not, in itself, say anything about whether the subsystem has its own dynamics. However, suppose we now consider a dynamical model in an N-particle sector, in which there is some sub-region R and some subset of $M < N$ particles which remain (over some time period) well inside R, while the other $N - M$ particles remain outside R; and suppose that as we approach the boundary ∂R of R, $\Phi(x,t)$ tends to some spatially-constant value $\Phi(\partial R, t)$. This could happen, for instance, because the other $N - M$ particles are very far from R, or because positive and negative charges outside R approximately balance so as to exert little net Couloumb force within R. Then to a good approximation, we can study the dynamics of the subsystem of those M particles autonomously from the other $N-M$: they will behave as if those particles were not present and as if the M particles were interacting among themselves according to the appropriate M-particle sector dynamics. If we want to study the dynamics of that subsystem of particles, in other words, under these assumptions and if we are sufficiently tolerant of small errors, we can replace the full N-particle sector with the M-particle sector. In this sense, the latter can be interpreted as modelling an isolated subsystem of the former.

Conversely, given the original N-particle sector, there will be many ways of embedding those N particles in a larger P-particle sector, with the remaining $P - N$ particles arranged so as to ensure the potential at the boundary of some region surrounding the N particles is approximately constant. In fact, Coulombic electrostatics has these two features:

(i) Any sector of the theory can be interpreted as an idealized description of an isolated subsystem of a sector of the theory;

(ii) An isolated subsystem of any sector of the theory can be described, in idealization, as a sector of the theory.

I will call a theory with these two features[7] *subsystem-recursive*: these are theories where any model can be interpreted *in the first instance* as modelling a

[7] They are independent. (i) might fail while (ii) holds in a theory whose sectors include a genuinely cosmological sector as well as sectors apt to model subsystems; for an example of the converse, see my discussion of Newtonian gravity with asymptotically constant boundary conditions, below.

dynamically isolated subsystem under certain idealizations about its environment and where, if we want to remove those idealizations, we can embed the model in a model of a larger system within the same theory — and where that larger system in turn is interpretable in the first instance as a subsystem of a still-larger system, with no assumption that we need to understand the whole setup in terms of an *ur*-system that describes the whole universe. The natural interpretation of a sector of a subsystem-recursive theory is always as a subsystem of some larger whole.

Before considering any more examples of subsystem-recursive theories, three subtleties are worth considering. Firstly, I have given no general definition of 'isolated' system. It's tempting to try something like 'subsystem where the causal influences of other particles are negligible' or 'subsystem where all the particles are very far from other particles'. But the first brings in metaphysical features that I'd prefer to avoid, and the second is clearly wrong in the electromagnetic case (it might be, e.g., that there are very many relatively-close particles, but positive and negative charges cancel out). A necessary condition of isolation is dynamical autonomy: that is, we can write down closed-form equations of motion for the subsystem that, to a good degree of approximation, govern the dynamics of its constituents, with the details of the rest of the system irrelevant except insofar as they do or do not satisfy the isolation condition. But beyond that, I think it's unlikely to be fruitful to look for a precise definition of 'isolated': a theory is subsystem-recursive if there is *some* robust, useful, category of dynamically-autonomous subsystems to which the definition applies.

The point can be sharpened by expanding Coulombic electrostatics to include theories with boundary conditions of the form

$$\lim_{|x|\to\infty} \Phi(x,t) = a_i(t)x^i + \Phi_0(t) \tag{4}$$

for arbitary time-dependent vector $a_i(t)$. Now, the class of 'isolated' systems can be taken to include subsets of particles in regions where the electric *field* (that is, the gradient of the potential) is constant on the boundary of the region (as would be appropriate in modelling the space between the plates of a capacitor, for instance). This expanded theory remains subsystem-recursive, and arguably is an improvement on the previous formulation of electrostatics inasmuch as it applies to a wider variety of dynamically-autonomous systems, but the charged particles in a strong electric field are not 'isolated' from that field in any *intuitive* sense of 'isolated'. (Nonetheless I will retain the term: in most of the cases I discuss in this paper, the intuitive meaning remains roughly correct.)

Secondly, there is some vagueness in my criterion for when a subsystem counts as isolated, and when it can be modelled by another sector of the theory. This is unavoidable: dynamical autonomy is by its nature relative to the desired degree of accuracy, the timescale over which the system is to be studied, and potentially the range of interventions to be considered. I want to suggest that this is just part of the normal modelling practice of physics: the idealized subsystem and the larger system both have fairly precise mathematical descriptions,

but the use of approximations in relating them is imprecise and to some extent an art as much as a science. Of course, if there is ever real doubt whether a subsystem is accurately modelled by its fully-isolated idealization, the question can be answered by solving the equations of motion for the full system — but by the time that has been done, the advantages of using the idealization are substantially diminished.

There is a close analogy here to the use of higher-level theories in physics (cf (Wallace 2012, ch.2) and references therein): there too, the higher-level and lower-level theories are typically both sharply defined in themselves, but the criteria under which the lower-level theory is said to instantiate the higher-level one are inexact and dependent on context and timescales. There too, we say that the higher-level theory only gets things right *ceteris paribus*, and the conditions for the validity of the ceteris paribus clause cannot be conclusively checked within the higher-level theory but require us to descend to the lower level.[8]

Thirdly (though relatedly) I say that a subsystem is represented by another *sector* of the theory, rather than just by some model in another sector, because it is crucial that the modal properties of the subsystem are represented: we care about the dynamical equations of the subsystem, not just about its actual occurrent features. But of course there will be a limit to which models in the sector actually succeed in representing the subsystem: if we consider states where the subsystem constituents lie well outside the original region R, so that they interact with the rest of the system in a way which violates the assumptions in the approximation, then those states no longer succeed in representing possible states of that subsystem — at least not *qua* isolated subsystem of the system. As with the issues of approximation we considered previously, it is inaccessible within a sector just when this occurs, because the external details that determine it have been idealized to infinity.

This, in turn, is a caveat for any would-be interpreter of a model. The details of that model should be taken as representing the physical goings-on of a system only relatively close to the action, so to speak. A model of some interacting particles, for instance, strictly speaking describes space outside the region occupied by those particles as empty, but it would be a mistake — a version of the Cosmological Assumption — to suppose that the theory really represents an infinity of empty space around the particles. In realistic applications of a physical theory, 'at infinity' does not mean 'literally infinitely far away': it means, more mundanely, something like 'well outside the region that the model is describing'. (A particularly striking example due to Porter Williams (in conversation): particle physicists analyse the particles produced by a collision process 'at infinity' — but the tube of the Large Hadron Collider is 3.8 meters across.[9])

This definition of subsystem recursivity is only preliminary: there is an intimate connection between subsystem-recursivity and symmetry, and so I now pause to develop the framework for symmetry that I use in this paper.

[8]See Reutlinger, Schurz, and Hütteman (2019) and references therein for extensive further discussion.

[9]Evans (2012, p.837).

3 Subsystem structure and symmetries

A key feature of the theories I discuss in this paper and its sequel is their rich subsystem structure, and it will be useful to discuss this at quite a high level of abstraction. Let me define a *subsystem structure* as a collection $\mathcal{X}$ of objects (the subsystems) equipped with a relation $\subset$ (where $X \subset Y$ should be read as 'X is a subsystem of Y'), obeying these assumptions:

- Partial ordering: $\subset$ is reflexive, transitive, and antisymmetric.
- Minimal element: there is some subsystem (written $\emptyset$) satisfying $\emptyset \subset X$ for all $X \in \mathcal{X}$.
- Maximal element: there is some subsystem (written $\mathcal{X}$ in a mild abuse of notation) satisfying $X \subset \mathcal{X}$ for all $X \in \mathcal{X}$.

Each sector of a particle theory can be interpreted as defining a subsystem structure: in an N-particle sector a subsystem is specified by a pair $X = \langle K, I \rangle$, where K is a subset of $\{1, \ldots N\}$ (identifying the particles to be included in the subsystem) and I is a subset of $\mathbb{R}$ (identifying the period of time for which the subsystem describes those particles). Given subsets $X_1 = \langle K_1, I_1 \rangle, X_2 = \langle K_2, I_2 \rangle$, $X_1 \subset X_2$ iff (i) X_2 describes all the particles that X_1 does, $K_1 \subset K_2$, and (ii) the set of times for X_1 is a subset of the set for X_2, $I_1 \subset I_2$.

A *state structure* for a subsystem structure $\mathcal{X}$ is then given by:

1. For each $X \in \mathcal{X}$, a set $S(X)$, the *state space* of X, satisfying $S(\emptyset) = \emptyset$.
2. For each $X, Y \in \mathcal{X}$ with $X \subset Y$, a map $r_{XY} : S(Y) \to S(X)$, the *restriction* of $S(Y)$ to $S(X)$, satisfying:
   - $r_{XX} = \mathrm{id}_X$, where id_X is just the identity on $S(X)$;
   - $r_{XY} \cdot r_{YZ} = r_{XZ}$.

   Given these conditions, the restriction from Y to X can be consistently written as $x \to x|_X$, since $(x|_Y)|_X = x|_X$.[10]

I impose one further requirement on this setup:

Upwards consistency: Given $X \subset Y$ and $x \in \mathcal{S}(X)$, there is some $y \in \mathcal{S}(Y)$ with $y|_X = x$. (Equivalently: each r_{XY} is onto, i. e. injective.)

(A brief note for readers familiar with the earlier framework for symmetries that I developed, jointly with Hilary Greaves, in Greaves and Wallace (2014): this framework differs (i) by allowing consideration of a plurality of different subsystems, rather than operating with a once-and-for all distinction between 'subsystem' and 'environment'; and (ii) by dropping the assumption that the total system ought to be thought of as representing the whole universe.)

[10] For readers who don't think I'm being abstract *enough*, here's an alternative definition: the subsystem structure $\mathcal{X}$ is a poset category with $\subset$ as its morphisms; a state structure is a *pre-sheaf* over $\mathcal{X}$, i. e. a contravariant functor from $\mathcal{X}$ to **Set**. (cf, e. g. , Awodey (2010)).

It will often be useful to talk about physical quantities for a subsystem: a quantity for X, by definition, is just a map from $\mathcal{S}(X)$ to some other space T. For instance, a *property* is defined by $T = \{0, 1\}$, or $T = \{\text{True}, \text{False}\}$ if you prefer. A real-valued quantity requires $T = \mathbb{R}$, and so on. Physical quantities for X do *not* restrict to subsystems of X, because the value of the quantity might depend on details of the state that are erased by restriction to X. They do have natural extensions from X to any Y with $X \subset Y$: the extension $\tilde{f}$ of f from X to Y is defined by $\tilde{f}(y) = f(y|_X)$.

In a particle theory, a state of the subsystem $\langle K, I\rangle$ is a smooth map from $K \times I$ to $\mathbb{R}^3$: put another way, it is an assignment, for each particle covered by the subsystem, of a path in $\mathbb{R}^3$ defined for all times in I. The restriction map is restriction in the ordinary functional sense. At this point, a note of terminology: subsystems in this context are *four-dimensional* entities, so that a 'state' of such a subsystem is a specification of its physical properties over an extended period of time. This notion of 'state' should be contrasted with the dynamical notion that arises when we talk about 'state space' and the like, and in which the dynamics of a system evolves one state into another. (In some ways 'history' would be clearer than 'state', but it is cumbersome.)

Continuing with the abstract presentation: Two subsystems X, Y are *disjoint* if $Z \subset X, Z \subset Y$ implies $Z = \emptyset$. Given subsystems X, Y, and states $x \in S(X)$, $y \in S(Y)$, there may or not be a state $z \in \mathcal{S}(\mathcal{X})$ with $Z|_X = x, Z|_Y = y$. If there is, I say that x and y are *compatible*. X and Y are *disconnected* if any $x \in S(X)$ and $y \in S(Y)$ are compatible. When subsystems are *contiguous* — that is, disjoint, but not disconnected — there are compatibility constraints on states in those subsystems. (Except in pathological cases, disconnectedness clearly implies disjointness, since x, y will not be compatible if their restrictions to a common subset of X and Y differ.)

In particle theory, for an example of disjoint but non-disconnected subsystems, take $X = \langle K, (0, 1)\rangle$ and $Y = \langle K, (-1, 0)\rangle$: the smoothness requirement on states requires states of the two subsystems to agree in a neighborhood of $t = 0$ to be compatible. (Conversely, it is easy to see that two particle-theory subsystems $\langle K_1, I_1\rangle$ and $\langle K_2, I_2\rangle$ are disconnected iff either K_1, K_2 are disjoint or the closures of I_1, I_2 are disjoint.)

A *symmetry group* for $\mathcal{X}$ is then an assignment, to each X, of a group $\mathcal{G}(X)$ in the algebraic sense, along with a homomorphism τ_{XY} from a subgroup $\mathcal{G}_X(Y) \subset \mathcal{G}(Y)$ to $\mathcal{G}(X)$. (The subgroup $\mathcal{G}_X(Y)$ should be interpreted as the group of symmetries of Y that restrict consistently to X; τ_{XY} is the restriction map.) The τ_{XY} are required to satisfy:

- $\tau_{XX} = \mathrm{id}_{\mathcal{G}(X)}$, where $\mathrm{id}_{\mathcal{G}(X)}$ is the identity homomorphism on $\mathcal{G}(X)$;
- Given $X \subset Y \subset Z$, then for any $g \in \mathcal{G}(Z)$ such that $\tau_{XY} \cdot \tau_{YZ} g$ is defined, $\tau_{XZ} g$ is also defined and $\tau_{XY} \cdot \tau_{YZ} g = \tau_{XZ} g$ (allowing us to consistently write the restriction of $g \in \mathcal{G}_X(Y)$ to $\mathcal{G}(X)$ as $g \to g|_X$);
- Each τ_{XY} is onto (injective), so that given $X \subset Y$ and $g \in \mathcal{G}(X)$, there is some $\tilde{g} \in \mathcal{G}(Y)$ with $\tilde{g}|_X = g$.

If $\mathcal{G}_X(Y) = \mathcal{G}(Y)$ for any $X \subset Y$, I say that the group *restricts fully*, i. e. all group elements preserve all subsystems.

An *action* of the symmetry group on a state structure is a group-theoretic action R_X of each $\mathcal{G}(X)$ on $S(X)$ such that for any X, Y with $X \subset Y$,

$$R_X(g|_X) = R_Y(g)|_X. \tag{5}$$

This has the consequence that restriction and group action commute, i. e. this diagram is commutative:

$$\begin{array}{ccc} x \in Y & \xrightarrow{g} & R_Y(g)x \\ {\scriptstyle r_{XY}}\downarrow & & \downarrow{\scriptstyle r_{XY}} \\ x|_X \in X & \xrightarrow{g|_X} & R_X(g|_X)x|_X \end{array}$$

In keeping with usual physics conventions, I will take the action of the group as tacit unless context demands otherwise, and, if $x \in S(X)$, will write gx for $R_X(g)x$. Given a group element $g \in \mathcal{G}(X)$, for any Y with $X \subset Y$ there will exist some (possibly non-unique) $\tilde{g} \in \mathcal{G}(Y)$ with $\tilde{g}|_X = g$; I call any such $\tilde{g}$ a *lift* of g to Y.

For any given $x \in X$, the *isotropy group* of x (sometimes called the *stabilizer group*) is the subset of $\mathcal{G}(X)$ whose elements leave x invariant. Normally speaking, the isotropy group of a state is trivial except for extremely special choices of state: for that reason, I define a state as *generic* if it has trivial isotropy group.

For particle theories, the physically-relevant symmetries come in two classes: *spacetime symmetries*, where some fixed (but possibly time-dependent) transformation is applied to each particle trajectory, and *permutation symmetries*, where some of the trajectories (in fact, always trajectories with the same masses and other parameters) are exchanged. The general form of a spacetime symmetry is

$$\vec{x}_I(t) \to F_t(\vec{x}_I(f(t))) \tag{6}$$

where F_t is a family of bijections of $\mathbb{R}^3$, and f is a bijection of $\mathbb{R}$; the general form of a permutation symmetry is

$$\vec{x}_I(t) \to \vec{x}_{\sigma(I)}(t) \tag{7}$$

where σ is some permutation of particle indices, normally required to satisfy some further constraints. For Coulombic particle theory, for instance, the spacetime symmetries have the form

$$\vec{x}_I(t) \to \mathbf{R}\vec{x}_I(\pm t - \tau) + \vec{s} + \vec{u}t \tag{8}$$

for arbitrary real τ, arbitrary vectors $\vec{s}$, $\vec{u}$, and arbitrary orthogonal matrix $\mathbf{R}$, and the permutation symmetries include all permutations between particles of the same masses and charges, as well as all permutations that swap particles with charge $+q$ to charge $-q$. Spacetime symmetries restrict fully: for any

subsystem of the theory, a spacetime symmetry restricts in a well-defined way to a bijection of the states of the subsystem. Permutation symmetries do not restrict fully: a permutation that permutes particles 2 and 3 does not restrict to a transformation on a set of particles containing particle 2 but not particle 3. However, given any subset of a set of particles, there is a well-defined subgroup of permutations that leaves the subset invariant, and any element of this subgroup defines a symmetry transformation on the subset.

4 Symmetry and discernibility for subsystems

Let's see how this framework can be used to interpret the symmetries of entire systems, via subsystem-recursivity. The basic *interpretative* assumption I make about these symmetries is that when two states of a *sub*system are related by a symmetry, they are identical in their *intrinsic* properties, so that any change caused by the symmetry at most changes a system's *extrinsic* properties, i. e. its relation to other systems. (And I assume that our description does not contain redundancy beyond that encoded in symmetries, so that the converse is true: any two states *not* related by a symmetry are differ in their intrinsic properties.) This reading of symmetry is fairly common (see Healey (2009) for an explicit statement of it) and I defend it at length, for dynamical symmetries, in Wallace (2019b); here I take it as a posit.

The framework we have developed now allows us to give a criterion for when a symmetry changes a system's extrinsic properties. For suppose X and Y are disjoint subsystems, with states x and y that are compatible: that is, such that some state z of the total system satisfies $z|_X = x$, $z|_Y = y$. And suppose that a symmetry $g \in \mathcal{G}(X)$ takes x to some $x' = gx$ which is also compatible with y, with some z' satisfying $z'|_X = x'$, $z'_Y = y$. Do any such z, z' differ by some intrinsic property? Yes — provided that there is *no* symmetry $\tilde{g} \in \mathcal{G}$ with $\tilde{g}z = z'$.

We can codify this into the following set of definitions: for any two $x, x' \in \mathcal{S}(X)$:

- x' is *intrinsically discernible* from x if there is no symmetry $g \in \mathcal{G}(X)$ such that $x' = gx$.
- x' is *extrinsically discernible* from x if there is some Y, and some $y \in \mathcal{S}(Y)$, such that any two $\tilde{x}, \tilde{x}'$ with $\tilde{x}|_X = x, \tilde{x}'|_X = x', \tilde{x}|_Y = \tilde{x}'|_Y = y$ are intrinsically discernible, i. e. are unrelated by any $g \in \mathcal{G}$. In this case, y is a *witness* to their extrinsic discernibility.
- x, x' are *indiscernible* if they are neither intrinsically nor extrinsically discernible

If a subsystem Y contains generic states (that is: states which are invariant only under the identity id_Y of $\mathcal{G}(Y)$), we can give a useful general criterion for when states of Y can witness the extrinsic discernibility of states $x, gx \in \mathcal{S}(X)$. Call g *Y-local* if it has a lift $\tilde{g}$ whose restriction to Y is the identity, $\tilde{g}|_Y = \mathrm{id}_Y$.

In this case, no state of Y can witness the discernibility of x, gx. Conversely, if g has no such lift, then provided X, Y are disconnected (so that any state of Y is guaranteed compatible with both x and gx), any generic state of Y will witness the discernibility of x, gx.

To apply this framework meaningfully, we need to require that the subsystem under consideration is small enough that we can regard the rest of the system as a reasonable proxy for all of its environment. For particle theory, we can consider a subsystem $\langle K, I\rangle$ to be small if (i) K contains a proper subset of all the particle labels; (ii) I is bounded. (So: a subsystem is small if it excludes some of the particles, and extends only over some finite time interval.)

We can also work at the level of quantities rather than states. A quantity f (recall, a quantity for X is a map from X to some target space T) is:

- *intrinsic* if whenever $x' = gx$ for $g \in \mathcal{G}(X)$, $f(x') = f(x)$; or, equivalently, if $f \cdot g = f$ for all $g \in \mathcal{G}(X)$.
- *extrinsic* if it is not intrinsic, but whenever x' and x are indiscernible, $f(x') = f(x)$.
- *gauge* if $f(x') \neq f(x)$ for x, x' indiscernible.

Where a theory includes gauge quantities, it is descriptively redundant: some of its states do not differ in any physically-relevant way and yet are mathematically represented as distinct. (Again, it only makes sense to apply this analysis directly for small subsystems.)

5 Symmetry and discernibility for general particle theories

Let's apply this framework to particle theories: given some subsystem $X = \langle K, I\rangle$ of a sector of a particle theory, and two states $x, gx \in \mathcal{S}(X)$ related by a symmetry $g \in \mathcal{G}(X)$, when are x and gx extrinsically discernible? There are two natural choices of witness to their extrinsic discernibility: either (i) some disjoint set of particles at the same time, or (ii) the same set of particles at a distinct time. These correspond respectively to (i) a subsystem $\langle K', I\rangle$ with $K \cap K' = \emptyset$, and (ii) a subsystem $\langle K, I'\rangle$ with I, I' having disjoint closure.

Following Wallace (2019b), a symmetry $g \in \mathcal{G}(\langle K, I\rangle)$ is *subsystem-local* iff it is Y-local for any $Y = \langle K', I\rangle$ with K, K' disjoint, and *time-local* iff it is Y-local for any $Y = \langle K, I'\rangle$ with I, I' having disjoint closure. If a symmetry is subsystem-local, symmetry-related states of the subsystem cannot be discerned by means of any other particle as a witness, but might be discernible using the same particles at an earlier or later time; if a symmetry is time-local, the converse is true.

Now: recall that a spacetime symmetry is specified by a possibly time-dependent bijection on space points (i. e., on $\mathbb{R}^3$) together with a bijection on

times (i. e. , on $\mathbb{R}$). Since each bijection is the same for all particles, no spacetime symmetry is subsystem-local; whether a spacetime symmetry is time-local depends on whether, and how, it depends on time. Permutation symmetries are subsystem-local (any permutation of a subset of particles is automatically a permutation of the whole set of particles) but not time-local (we cannot permute the particles at one time but not at another, not if we wish the transformation to be a dynamical symmetry).

As a consequence, any two states related by a (non-trivial) spacetime symmetry are extrinsically discernible, with any disjoint collection of particles serving as a witness to the discernibility; any two states related by a (non-trivial) permutation symmetry are likewise extrinsically discernible, with those same particles at an earlier or later time witnessing it. In the case of the Coulomb theory we have mostly been discussing so far (whose spacetime symmetry group is the Galilei group), the spacetime symmetries are also time-local, so that specifying the symmetry for some set of particles over any time interval suffices to specify it for that same set over all other time intervals — hence, a particle's earlier state also serves to witness the discernibility of symmetry-related states. (In section 6 we will see that this is not always the case for spacetime symmetries.)

So: states differing by any particle-theory symmetry — whether spacetime, permutation, or a combination thereof — are extrinsically discernible: there are no distinct-but-indiscernible states in particle mechanics. At the level of quantities: all particle-theory quantities are either intrinsic or extrinsic: there are no gauge quantities in particle mechanics, and hence no descriptive redundancy.

The core idea of subsystem recursivity is that we can now apply these conclusions to entire sectors of a theory, because those sectors should be interpreted as describing subsystems of larger systems under the idealization that the rest of the system can be neglected. Doing so, we reach the conclusion that all symmetries in particle theory should be understood as physical transformations, taking a model to an indiscernible-yet-distinct model.

It is useful to see how this plays out for the dynamical conception of states (where states are phase-space points, i. e. specification of all particle positions and momenta, rather than segments of particle histories). Formally we can define a subsystem structure on N-particle phase space, with the subsystems just correspending to subsets of the particles. Under this analysis, spacetime symmetries of a subsystem X are never X-local, but permutation symmetries of the subsystem always are.

That gives us good reason to regard spacetime-symmetry-related dynamical states as intrinsically identical but extrinsically different. The reason has a direct physical interpretation: the application of a spacetime symmetry to a subset of the particle states changes their positions and/or velocities relative to other particles. Given subsystem recursivity, this should be carried over to symmetries of the whole system, in which those 'other particles' are idealized to spatial infinity but not obliterated.

In this context it is interesting to consider the periodic suggestion that in the presence of a symmetry, the 'true' state space is obtained by quotienting under the action of any symmetries. This can be done in particle mechanics

(the general process is known as dynamical reduction, or as Marsden-Weinstein reduction; see (Marsden and Ratiu 1999) and references therein for more details; see also Wallace (2019b) for discussion) and does indeed yield a well-defined *intrinsic* dynamics for the system — but that dynamics does not encode the change over time of the system's extrinsic features, i. e. how it changes relative to the background of other systems. So for that reason, quotienting is throwing away real information (at least as long as we reject the Cosmological Assumption in our interpretative practice).

What about the permutation symmetries? Here I think it is a matter of convention: do we want to define the state space of the theory such that it encodes the extrinsic permutation-variant properties, or not? As long as we regard the history state as fundamental and the dynamical state as derivative, this seems to be up to us: the quotiented description has a certain austere clarity, against which the unquotiented description is usually easier to work with. But nothing factual hangs on this choice.

6 Symmetry and subsytem recursivity for different versions of Newtonian gravity

To explore both this symmetry-based analysis of intrinsic and extrinsic quantities, and the general idea of subsystem recursivity, in this section I want to consider a number of different formulations of Newtonian gravitation that have been discussed in the recent literature. Each of these theories fits the general particle-mechanics mold: each sector has a fixed finite number N of particles, and a fixed assignment of masses to the particles; each model is specified by a trajectory $\vec{x}_I(t)$ for each particle I. For simplicity, I disregard permutation symmetries in the remainder of the paper (they work out the same in each case) and focus entirely on the spacetime symmetries.

The first example is *Newtonian potential gravity*. In this theory, sectors are individuated by particle number and masses, but also by a spatial boundary condition for a real field Φ on $\mathbb{R}^3 \times \mathbb{R}$, which for the moment I will require to take the form (previously seen in eqn. ((4)):

$$\lim_{|x|\to\infty} \Phi(x,t) = a_i(t)x^i + \Phi_0(t) \tag{9}$$

for some time-dependent vector $a_i(t)$. In dynamical models, the trajectories satisfy

$$\frac{\mathrm{d}}{\mathrm{d}t} x^i_J(t) = \nabla_i \Phi(x^i_J(t), t) \tag{10}$$

where $\Phi(x,t)$ is the unique solution to the equation

$$\nabla^2 \Phi(x,t) = -4\pi G \sum_J \delta(x - x_J) m_J \tag{11}$$

subject to the boundary condition (9). (As usual G is fixed and sector-independent; it could be set to 1 by an appropriate choice of units.) Subsystems are defined

just as in the electrostatic theory, but in Newtonian gravity, the *isolated* subsystems are freely falling subsystems in which tidal forces can be neglected, i. e. subsystems where the potential approximates the form $a_i(t)x^i + \Phi_0(t)$ on the boundary for given $a_i(t), \Phi_0(t)$.

(As with the Coulomb case, it is possible to relax the boundary conditions to permit a wider class of subsystem. We might, for instance, permit boundary conditions of the form

$$\lim_{|x|\to\infty} \Phi(x,t) = M_{ij}(t)x^i x^j + a_i(t)x^i + \Phi_0(t) \tag{12}$$

in which case the specific boundary condition

$$\lim_{|x|\to\infty} \Phi(x,t) = -\frac{GM}{r_0} + \frac{3GM}{4r_0^3}\left(r_0^2 x_i x^i - 3(r_{0i}(t)x^i)^2\right) \tag{13}$$

for $\vec{r}_0 = (r_0 \cos\omega t, r_0 \sin\omega t, 0)$ and $\omega^2 = GM/r_0$ describes physics in the vicinity of a point in circular orbit distance r_0 from a distant mass M. Plausibly more general quadratic boundary conditions can also be interpreted as subsystems — but for my purposes the simpler condition (11) suffices.)

The Galilean symmetries are symmetries for this theory, and indeed are the full set of *sector-preserving* symmetries, but there is a larger class of symmetries that relate states in different sectors. Specifically, consider transformations of form

$$\vec{x}_n(t) \to \mathbf{R}\vec{x}_n(t-\tau) + \vec{s}(t) \tag{14}$$

together with changes of boundary condition of form

$$\vec{a}(t) \to \vec{a}(t) - \ddot{\vec{s}}(t) \tag{15}$$

for arbitrary twice-differentiable vector function $\vec{s}(t)$. If $\ddot{\vec{s}}(t) \neq 0$, this corresponds to a time-dependent acceleration imposed on all bodies, together with a time-dependent, spatially-constant gravitational field sufficient to generate that acceleration. Let's call this larger group of symmetries the *Maxwell group*, following Earman (1989). It includes the Galilean transformations in the special case where $\ddot{\vec{s}}(t) = 0$, and in that case the boundary condition is invariant. In other cases, it is transformed, but the transformation remains a dynamical symmetry.

It follows that purely spatial-translation symmetries (i. e. , those with $\mathbf{R} = \mathrm{id}$ and $\tau = 0$) are time-local: their form over one time interval does not fix their form over another time interval, and it does not directly make sense to talk about where a particle is relative to its own earlier location. However, these symmetries are not subsystem-local, so that we can regard the symmetries as changing the extrinsic position and velocity of the system (understood, as usual, as defined relative to the backdrop of other subsystems). Note also that despite the presence of a time-dependent symmetry, the theory is deterministic: time-dependent transformations (other than the Galilean ones) map between different sectors, but dynamics is defined within a sector.

Let's pause briefly to consider the direct empirical significance of these various symmetries (cf Kosso 2000, Brading and Brown 2004, Greaves and Wallace 2014). The Galilean symmetries, since they leave the sector invariant, can be understood as transformations of a subsystem that leave the rest of the system unchanged (at least in the idealization where dynamical interaction between subsystems is neglected). The non-Galilean transformations cannot be so understood: they change the boundary conditions, and so some change in the surrounding subsystems is required. In this particular case, doing so requires a modification of the gravitational field in which the subsystem is located. But the transformation can still be understood in a well-defined way as altering the extrinsic quantities of the subsystem, because we can use an arbitrary (and indeed arbitrarily distant) other subsystem as the reference with respect to which those quantities are to be interpreted. Indeed, the existence of this symmetry is familiar to us: it is simply Einstein's equivalence principle, stating that an external gravitational field has no effect on the intrinsic behavior of a system.

In Greaves and Wallace *ibid*, we distinguished two classes of scenario that gave rise to direct empirical significance: *Galileo's ship scenarios*, in which the symmetry could be understood as acting on a subsystem while leaving the rest of the system unchanged, and *Einstein's elevator scenarios*, in which applying the symmetry also required changes to the system's surroundings. At least in the context of particle mechanics (the situation is subtler in field theory, as I discuss in Part II), we can give clean conditions for when these arise: in either case the symmetry must not be be subsystem-local (so that reference systems exist to witness the change); if such a symmetry is sector-preserving, it gives rise to Galileo's-ship scenarios; if it is not sector-preserving, it gives rise to Einstein's-elevator scenarios.

We can also identify an alternative form of direct empirical significance, not much considered in this literature, which occurs when a symmetry is not *time*-local: in that case, we can consider the same system before and after the symmetry transformation is applied to it by some external process (e.g. by swapping the particles around or by physically rotating the system) and observe that this physically meaningful change nonetheless makes no difference to the system's intrinsic dynamics. (There is actually a reasonable argument that the historical Galileo's-ship thought experiment ought to be understood this way, since after all Galileo does speak of putting the ship into motion; still, the norm in the current literature is to consider the application of the symmetry counterfactually rather than temporally.)

For our second example of a Newtonian gravity theory, let's suppose that we simply restrict our attention to boundary conditions that are asymptotically constant. This restricts the symmetry group down to the Galilean group and so demotes the equivalence principle from being a *symmetry* principle. (An essentially equivalent theory is obtained by eschewing use of a potential and simply working with an inverse-square force law: the inverse-square force is, provably, the gradient of the unique solution to Poisson's equation with asymptotically constant boundary conditions.)

What (if anything) is wrong with this restriction? Knox (2014) has argued

that it essentially corresponds to an arbitrary and unobservable split between gravitational and inertial structure, and that seems right to me (see Wallace (2020a) for further discussion), but the present framework lets us make the point more directly: in introducing the restriction on the boundary conditions, we break the subsystem-recursivity of the theory. For freely-falling systems in regions devoid of tidal forces are isolated subsystems, and yet they are not modelled by any sector of the theory.

The careful reader may object: did I not claim that 'isolated subsystem' was a term of art, not something with a fixed pre-theoretic meaning? Indeed so. And we could try defining a system as 'isolated' if it is moving at constant speed with respect to the background Galilean geometry defined by the fixed boundary condition. Knox's concern about unobservability is now pressing, and it too can be captured within the present framework: given an isolated system in Newtonian mechanics according to our current definition, there are other possible states of affairs where that system retains the same intrinsic dynamics while ceasing to be 'isolated'. A reasonable condition on which systems count as isolated is that if system X counts, and system Y is intrinsically identical to system X, then system Y should also count; it is this condition that is violated by our restriction on boundary conditions. (This 'reasonable condition' requires the notion of intrinsic discernibility to be stated: here, as advertised earlier, we see the interplay between subsystem-recursivity and symmetry.)

How serious is it for our theory not to be subsystem-recursive? The problem is not that the theory is metaphysically unacceptable. Rather, it is that we want to use the theory to study isolated subsystems of the universe (it is not an empirically plausible cosmology, after all!) And trying to do so is unstable: once we have established that inertially-moving systems are to be treated as isolated, we quickly recognize that freely-falling systems also need to be treated as isolated, since they are intrinsically indiscernible from the original class of isolated systems. That leads us to expand the modelling capacity of our theory by allowing more general boundary conditions; that in turn expands the symmetry group, and so we learn that we were wrong about the symmetries of our theory. Not wrong in the sense that a hypothetical theory of cosmology *could not* have that symmetry group, but wrong in the sense that the symmetries of subsystems of the theory's models, according to the theory itself, are larger than the symmetries of the original theory.

Something similar happens in two other versions of Newtonian gravity that have received philosophical attention, but there, from the present perspective, the symmetry group is too large rather than too small. In the relational dynamics developed by Barbour and co-workers (Barbour and Bertotti (1982), Barbour (1982); see Belot (1999), Pooley and Brown (2002), Barbour (2012) and references therein for detailed discussion), the symmetry group is

$$\vec{x}_n(t) \to \mathbf{R}(t)\vec{x}_n(f(t)) + \vec{s}(t) \tag{16}$$

for an arbitrary time-dependent orthogonal matrix $\mathrm{R}(t)$, an arbitrary smooth bijection $f : \mathbb{R} \to \mathbb{R}$, and an arbitrary time-dependent translation $\vec{s}(t)$, and the

sectors are individuated by particle number and mass, with no boundary conditions to consider. The actual form of the dynamics for the theory is somewhat complicated to state and I refer readers to the references given, but it can be summarised for our purposes by saying that a model is dynamically allowed iff it is symmetry-related to a solution of 'ordinary (i. e. Newtonian-potential) gravity with total energy and angular momentum equal to zero.[11]

This theory fails to model its own subsystems. For given an isolated subset of particles, their intrinsic dynamics will be that of N particles in Newtonian gravitation, generically with a *non*-zero angular momentum and energy, so that the state space of the subsystem contains states not present in Barbour's theory. It does not suffice just to declare by fiat that a subsystem is isolated only if its energy and angular momentum vanish: we need to get the modal structure of the subsystem right, and even if it *in fact* has zero angular momentum, it perfectly well *could have* had nonzero angular momentum. Furthermore, even for a subsystem with zero angular momentum the rotational and translational motion of the subsystem has (extrinsic) physical meaning, but is not tracked by the relational dynamics.

For this reason, Barbour's relationism is *not* subsystem-recursive, and we cannot interpret it in that light. It is an *intrinsically cosmological* theory, which can be understood only as a model of the whole universe. Barbour himself recognizes this, of course:

> [L]ocally the predictions ... do not differ at all from conventional Newtonian dynamics. It is only if one has access to the entire universe that experimental verification of the Leibnizian predictions of vanishing energy and angular momentum for the universe as a whole becomes possible. (Barbour 1982, p.273)

If one were to adopt relational dynamics as one's theory of the universe as a whole, one would still end up with ordinary Newtonian gravitation as the correct (and subsystem-recursive) *non*-cosmological theory. Since in the actual universe relational dynamics is not empirically adequate as a cosmology (no obvious generalisation to infinitely many particles, no dark energy, no asymptotically de Sitter geometry, etc., it follows (*contra* Pooley and Brown (2002)) that the scientific value of relational dynamics is as a thought experiment, or a demonstration of the logical coherence of Machian ideas, or a historical what-if. Unlike Newtonian gravitation (which of course is widely used to this day in physics, to model systems ranging in scale from lab experiments to galaxies), relational dynamics does not appear to have a role modelling actually-existing physical systems.

Somewhat similar conclusions can be drawn for our final gravitational theory: Saunders' *vector relationism* (Saunders (2013), though the name is coined in (Wallace 2020a)). Here the symmetry group is the same as for Newtonian potential gravity (arbitrary time-dependent spatial translations, combined with

[11] It is straightforward to generalise the theory so that the energy takes on some fixed and sector-independent, but non-zero, value.

time-independent rotations and time translations) but the sectors are individuated as in Barbour's theory, by particle masses and numbers alone. Vector relationship, therefore, can be expected to be indeterministic (we can apply symmetries with arbitrary time dependence whilst remaining in a sector, just as for Barbour's relationism), and indeed this can be read off directly from its dynamics.

Saunders' theory is also not subsystem-recursive (as I spell out in rather more detail in Wallace (2020a)): its isolated subsystems have center-of-mass accelerations that are determined entirely by the matter distribution in other subsystems, and so must be invariant under symmetries, and this breaks the symmetry group of subsystems down to the Galilei group. (It could be made subsystem-recursive by a finer-grained choice of sector (taking advantage of the definitional requirement that dynamically-permissible models remain within a sector), but this would effectively reduce it to Newtonian potential gravity.) As such, the theory is intrinsically cosmological (as Saunders, again, recognizes): in non-cosmological applications, we are led from vector relationism to the Newtonian potential theory. There is, however, another interpretation of Saunders' theory: as the dynamical reduction of Newtonian gravity. For the theory's models do correctly describe the *intrinsic* dynamics of isolated subsystems. (This is in contrast to Barbour's theory, which fails to do so owing to the stipulations that energy and angular momentum must vanish; no analogous stipulation holds for Saunders.)

To conclude this discussion: subsystem recursivity is a powerful criterion to select between different versions of Newtonian gravity which, if regarded as cosmological, are empirically equivalent or nearly so. But the reasons for applying that criterion are not epistemic or metaphysical, but methodological: even if we were to choose a subsystem-non-recursive theory to model the whole universe, we will end up moving to the subsystem-recursive theory as long as we want to study isolated subsystems of the universe — and in practice, the application of the theory to these subsystems will pretty much exhaust both the scientific evidence for, and the scientific usefulness of, the theory.

7 Conclusions

In physical practice, virtually all theories are used to model isolated subsystems of a larger universe (and, normally, to model even those systems only at a certain level of analysis); they are hardly ever used as models of the universe as a whole. In many cases, though, including most of the wide-ranging theories used by physicists, theories are subsystem-recursive: their models can incorporate isolated proper subsystems which are modelled by other sectors of the same theory, and can in turn be understood as modelling isolated subsystems of larger systems also modellable by other sectors of the theory.

Subsystem recursivity offers a general framework to assess the epistemic, modal and metaphysical status of the symmetry transformations of a theory, by firstly considering symmetry transformations restricted to some subsystem of a

sector and then translating the results to sectors as a whole via subsystem recursivity. When this general framework is applied to the spacetime symmetries of nonrelativistic particle theories, the result is that any such symmetry, applied to a subsystem, cannot be extended to a symmetry of the whole system that leaves any other system invariant, and so spacetime symmetry transformations give rise to extrinsically distinct and empirically distinguishable states of affairs. Applied to the permutation symmetries of those theories, the result is subtler: such transformations, applied to a subset of particles, can be so extended to any other particles, and so the discernibility of states related by such a transformation cannot be witnessed by any other particles — but it can be witnessed by earlier or later states of the same particles.

Subsystem recursivity has implications for the interpretation of symmetries, and in turn a clear understanding of symmetries allows a cleaner statement of subsystem recursivity, clarifying what it is for two subsystems to be intrinsically indiscernible and so allowing us to insist that a subsystem-recursive theory, if it treats a state of a system as isolated, must treat intrinsially indiscernible states of that same system as isolated. So understood, subsystem recursivity has uses beyond the interpretation of symmetry, as I illustrated through an examination of various forms of Newtonian gravity: the standard potential-theory version is subsystem-recursive, but various alternatives — inverse-square-law gravity, the Barbour-Bertotti theory, and Saunders' vector relationism — for various reasons are not. This provides a methodological argument in favor of the potential-theory form, at least for the normal purposes for which Newtonian gravity is used.

I hope to have demonstrated in this paper the interpretative values of this approach to isolated systems and their symmetries, but of course my examples remain restricted to relatively simple physical theories. In the sequel I extend them to the more complicated situations which arise in field theory.

Acknowledgements

I am grateful to Gordon Belot, Harvey Brown, Neil Dewar, Eleanor Knox, Oliver Pooley, Laura Ruetsche, Simon Saunders, Nic Teh, and Jim Weatherall for the many conversations on symmetry which led to this paper. It also benefitted significantly from feedback at meetings in Louvain-le-Neuve and Hannover in 2017, and in Salzburg in 2019. Thanks also to Valeriya Chasova and two anonymous referees for helpful comments on a previous draft.

References

Albert, D. Z. (1996). Elementary quantum metaphysics. In J. T. Cushing, A. Fine, and S. Goldstein (Eds.), *Bohmian Mechanics and Quantum Theory: An Appraisal*, Dordrecht, pp. 277–284. Kluwer Academic Publishers.

Allori, V., S. Goldstein, R. Tumulka, and N. Zanghí (2008). On the common

structure of Bohmian mechanics and the Ghirardi-Rimini-Weber theory. *British Journal for the Philosophy of Science 59*, 353–389.

Awodey, S. (2010). *Category Theory*. Oxford: Oxford University Press.

Barbour, J. (1982). Relational concepts of space and time. *British Journal for the Philosophy of Science 33*, 251.

Barbour, J. (2012). Shape dynamics: an introduction. In F. Finster, O. Müller, M. Nardmann, J. Tolksdorf, and E. Zeidler (Eds.), *Quantum Field Theory and Gravity: Conceptual and Mathematical Advances in the Search for a Unified Framework*, pp. 257–297. Springer. Available online at http://arxiv.org/abs/1105.0183.

Barbour, J. and B. Bertotti (1982). Mach's principle and the structure of dynamical theories. *Proceedings of the Royal Society of London 382*, 295.

Belot, G. (1998). Understanding electromagnetism. *British Journal for the Philosophy of Science 49*, 531–555.

Belot, G. (1999). Rehabilitating relationism. *International Studies in the Philosophy of Science 13*, 35–52.

Belot, G. (2011). Background-independence. *General Relativity and Gravitation 43*, 2865–2884.

Belot, G. (2013). Symmetry and equivalence. In R. Batterman (Ed.), *The Oxford Handbook of Philosophy of Physics*, pp. pp. 318–339. Oxford: Oxford University Press.

Belot, G. (2018). Fifty million Elvis fans can't be wrong. *Nous 52*, 946–981.

Brading, K. and H. Brown (2004). Are gauge symmetry transformations observable? *British Journal for the Philosophy of Science 55*, 645–665.

Brown, H. and R. Sypel (1995). On the meaning of the relativity principle and other symmetries. *International Studies in the Philosophy of Science 9*, 235–253.

Dasgupta, S. (2016). Symmetry as an epistemic notion (twice over). *British Journal for the Philosophy of Science 67*, 837–878.

Earman, J. (1989). *World Enough and Space-Time*. Cambridge, Massachusetts: MIT Press.

Earman, J. (2004). Laws, symmetry, and symmetry breaking: Invariance, conservation principles, and objectivity. *Philosophy of Science 71*, 1227–1241.

Evans, L. (2012). The Large Hadron Collider. *Philosophical Transactions of the Royal Society A 370*, 831–858.

Feynman, R. P. and A. R. Hibbs (1965). *Quantum Mechanics and Path Integrals*. New York: McGraw-Hill.

Greaves, H. and D. Wallace (2014). Empirical consequences of symmetries. *British Journal for the Philosophy of Science 65*, 59–89.

Healey, R. (2009). Perfect symmetries. *British Journal for the Philosophy of Science 60*, 697–720.

Ismael, J. and B. van Fraassen (2003). Symmetry as a guide to superfluous theoretical structure. In K. Brading and E. Castellani (Eds.), *Symmetries in Physics: Philosophical Reflections*, pp. 371–392. Cambridge: Cambridge University Press.

Knox, E. (2014). Newtonian spacetime structure in light of the equivalence principle. *British Journal for the Philosophy of Science 65*, 863–880.

Kosso, P. (2000). The empirical status of symmetries in physics. *British Journal for the Philosophy of Science 51*, 81–98.

Marsden, J. E. and T. S. Ratiu (1999). *Introduction to Mechanics and Symmetry*. Springer-Verlag.

Møller-Nielsen, T. (2017). Invariance, interpretation, and motivation. *Philosophy of Science 84*, 1253–1264.

Ney, A. and D. Albert (Eds.) (2013). *The Wave Function: Essays on the Metaphysics of Quantum Mechanics*. Oxford: Oxford University Press.

Pooley, O. and H. R. Brown (2002). Relationism rehabilitated? i: Classical mechanics. *British Journal for the Philosophy of Science 53*, 183–204.

Reutlinger, A., G. Schurz, and A. Hütteman (2019). *Ceteris Paribus* laws. In Edward N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2019 edition). Available at https://plato.stanford.edu/archives/spr2019/entries/ceteris-paribus/.

Ruetsche, L. (2011). *Interpreting Quantum Theories*. Oxford: Oxford University Press.

Saunders, S. (2003). Physics and Leibniz's principles. In K. Brading and E. Castellani (Eds.), *Symmetries in Physics: Philosophical Reflections*, pp. 289–308. Cambridge: Cambridge University Press.

Saunders, S. (2013). Rethinking Newton's *Principia*. *Philosophy of Science 80*, 22–48.

Van Fraassen, B. C. (1980). *The Scientific Image*. Oxford: Oxford University Press.

Wallace, D. (2012). *The Emergent Multiverse: Quantum Theory according to the Everett Interpretation*. Oxford: Oxford University Press.

Wallace, D. (2019a). Isolated systems and their symmetries, part II: Local and global symmetries of field theories. *Studies in the History and Philosophy of Science*, this volume.

Wallace, D. (2019b). Observability, redundancy and modality for dynamical symmetry transformations. Forthcoming in J. Read, B. Roberts and N. Teh (eds.) *The Philosophy and Physics of Noether's Theorem* (CUP, forthcoming). Preprint online at http://philsci-archive.pitt.edu/18813/.

Wallace, D. (2020a). Fundamental and emergent geometry in Newtonian dynamics. *British Journal for the Philosophy of Science 71*, 1–32.

Wallace, D. (2020b). On the plurality of quantum theories: Quantum theory as a framework, and its implications for the quantum measurement problem. In J. Saatsi and S. French (Eds.), *Scientific Realism and the Quantum*, pp. 78–102. Oxford: Oxford University Press.

Williams, P. (2019). Scientific realism made effective. *British Journal for the Philosophy of Science 70*, 209–237.