(Competing?) Formulations of Newtonian Gravitation: Reflections at the Intersection of Interpretation, Methodology, and Equivalence

Forthcoming in the *Journal of Philosophy*

Kevin Coffey*
Department of Philosophy
New York University Abu Dhabi

Abstract

It is sometimes said there are two ways of formulating Newtonian gravitation theory. On the first, matter gives rise to a gravitational field deflecting bodies from inertial motion within flat spacetime. On the second, matter's accelerative effects are encoded in dynamical spacetime structure exhibiting curvature and the field is 'geometrized away'. Are these two accounts of Newtonian gravitation theoretically equivalent? Conventional wisdom within the philosophy of physics is that they are, and recently several philosophers have made this claim explicit. In this paper I develop an alternative approach to Newtonian gravitation on which the equivalence claim fails, and in the process identify an important but largely overlooked consideration for interpreting physical theories. I then apply this analysis to (a) put limits on the uses of Newtonian gravitation within the methodology of science, and (b) defend the interpretive approach to theoretical equivalence against formal approaches, including the recently popular criterion of categorical equivalence.

*Thanks to Pablo Acuña, Dave Baker, Gordon Belot, Hartry Field, Josh Hunt (especially), Michel Janssen, and Tim Maudlin, and to audiences at New York University, the University of Michigan, and the Second Chilean Conference on the Philosophy of Physics. Thanks also to my Orient Pandemic Crew of 2020: Jen, Michael, and Dominick.

1 Introduction

It is sometimes said that there are two modern ways of formulating Newtonian gravitation theory. The first, which is what usually falls under the label *Newtonian gravitation theory* (NG), models gravitating systems against the backdrop of a flat, 4-dimensional manifold representing spacetime. The remaining mathematical objects and equations invoked in this representation have suggested to many people a physical picture according to which matter gives rise to a gravitational field, the strength and direction of which causes bodies to accelerate and thereby be deflected from inertial motion. The fixed spacetime background plays a central role here, grounding the distinction between inertial motion (force-free unaccelerated motion in a straight line) and the effects of gravitational interactions. This basic picture is ubiquitous in textbooks on classical dynamics.[1]

There is also a second, lesser-known formulation of Newtonian gravitation theory, initially developed using the tools of modern differential geometry in the wake of Einstein's general theory of relativity: *geometrized Newtonian gravitation theory* (GNG).[2] What would otherwise be characterized (on the standard understanding) as the accelerative effects of the gravitational field are now encoded in a smooth manifold exhibiting curvature, taken to represent a spacetime possessing dynamical geometrical structure. As in General Relativity, spacetime curvature depends on the overall distribution of matter in the world. Massive bodies are no longer understood as being deflected from inertial motions by a gravitational field, but instead

[1]See, e.g., Marion and Thornton (1995) and Taylor (2005), although neither textbook formulates things explicitly in terms of spacetime.

[2]This formulation is often called *Newton-Cartan theory*. Good sources include Havas (1964), Trautman (1965), Malament (1986), and Malament (2012).

follow geodesic ('straightest') trajectories in a curved spacetime in which the gravitational field has been 'geometrized away'. Those trajectories are just taken to *be* the inertial motions.

The obvious differences in the physical pictures presented above might leave one puzzled regarding talk of different 'formulations' of Newtonian gravitation theory. Do we not have here two entirely distinct physical theories? A growing number of philosophers and physicists have endorsed or expressed sympathy for the claim that NG and GNG are but reformulations of each other—that some part of the standard physical glosses I've given to the formulations above is in fact grossly misleading or defective—and that they ought to be understood as theoretically equivalent. This equivalence claim is intended in much the same way that formulations of classical electrodynamics in terms of fields and potentials are generally thought to be theoretically equivalent. Although it is *possible* to understand the different formulations as positing distinct ontologies obeying distinct dynamical laws, it is widely agreed that this is not how they are *best* understood. The fields are generally taken to be physically real, whereas the potentials are taken to be mathematical artifacts. In this sense, not only are the field- and potential-based formulations theoretically equivalent, but it is the former and not the latter that is 'ontologically perspicuous'. So, too, in the case of Newtonian gravitation: proponents of the equivalence claim have by-and-large identified GNG as the ontologically perspicuous formulation. Those features of NG that suggest a real gravitational field existing in a fixed flat spacetime are, like the electrodynamical potentials, taken to be mere mathematical artifacts.

This paper argues that there is a justifiable and philosophically relevant approach to NG on which NG and GNG are not theoretically equivalent. In this way it aims to challenge the equivalence claim. On the view I develop the formulations ought to be understood as distinct but observationally equivalent rival theories. After a brief technical overview I present the central argument for theoretical equivalence, and then motivate a way of thinking about NG on which the argument fails. This approach makes salient an interpretive consideration regarding inter-relationships between formalisms that has been largely overlooked in the philosophy of physics literature. I then use the preceding discussion to draw two further lessons. First, I argue that there are limits on the ways that NG can plausibly be used within the methodology of science. Second, I defend the interpretive approach to theoretical equivalence against formal approaches, including the recently popular criterion of categorical equivalence.

2 Newtonian Gravitation Two Ways?

This section provides a brief overview of the model-theoretic structures of NG and GNG in order to highlight the mathematical relationships that figure centrally in the argument for theoretical equivalence. GNG is an inherently 4-dimensional theory, involving as it does the curvature of spacetime, and so the starting point for our discussion is the notion of a classical spacetime model.

2.1 Classical Spacetime Models

Following Malament (1986) and Malament (2012), I define the notion of a classical spacetime model as follows:

Definition 1 *A* ***classical spacetime model*** *is a structure* $(M, h^{ab}, t_{ab}, \nabla_a)$*, where:*

1. *M is a smooth, (simply) connected differentiable manifold;*
2. *h^{ab} is a smooth, nonvanishing symmetric $(2,0)$-tensor field on M of degenerate signature $(0,1,1,1)$;*
3. *t_{ab} is a smooth, nonvanishing $(0,2)$-tensor field on M of degenerate signature $(1,0,0,0)$;*
4. *∇_a is a smooth covariant derivative operator (or affine connection) on M with associated Christoffel symbols $\Gamma^{\alpha}{}_{\beta\nu}$;*
5. *$h^{ab}t_{bc} = 0$ (orthogonality); and*
6. *$\nabla_a t_{bc} = \nabla_a h^{bc} = 0$ (compatibility).*

Conceptually, M represents spacetime; h^{ab} and t_{ab} encode the spatial and temporal metrics, respectively; and ∇_a defines a notion of *constancy* or *parallel transport* between tensors at different points. Conditions 1–6 together guarantee that this structure characterizes a spacetime that can (among other things) be foliated into hypersurfaces representing flat 3-dimensional space at distinct times, although no means is provided for identifying spatial points at distinct times. Nothing in the definition requires that the spacetime itself—as opposed to the spatial hypersurfaces—be flat.[3]

A particle trajectory is represented by a *time-like* curve in M, which is a curve γ such that $t_{ab}\xi^a\xi^b > 0$, where ξ^a is the tangent vector field along γ. Intuitively, the trajectory of a particle must always have a non-zero temporal component. A particle is accelerating at a point p along its trajectory iff $\xi^a\nabla_a\xi^b \neq 0$ at p, so time-like curves for which $\xi^a\nabla_a\xi^b = 0$ correspond to possible trajectories of inertially moving particles. In this way, ∇_a encodes the inertial structure of the spacetime and provides the standard for distinguishing inertial motions from accelerated ones. But unlike the

[3]How conditions 1–6 give rise to this interpretation is discussed in Earman and Friedman (1973), Malament (1986), Malament (2007), Malament (2012), Stachel (2007), and Trautman (1965). Throughout I omit proofs and many technical details, but refer the reader to Malament (2012, ch.4) for an especially clear discussion. As a notational matter, for any one-form s_a in what follows, s^a is an abbreviation for $h^{ab}s_b$.

case of a manifold equipped with a non-degenerate metric, as postulated in relativity theory, the choice of an affine connection here is not uniquely determined by the spatial and temporal metrics. By selecting a different affine connection on M satisfying compatibility—say, $\nabla'_a \neq \nabla_a$—we arrive at a different classical spacetime model $(M, h^{ab}, t_{ab}, \nabla'_a)$ endowed with a different standard for distinguishing inertial and accelerated motions.

2.2 NG-Models

Given the notion of a classical spacetime, a model of NG can then be defined as follows:[4]

Definition 2 *An* ***NG-model*** *is a structure $(M, h^{ab}, t_{ab}, \nabla_a, \phi, \rho)$, where:*

1. *$(M, h^{ab}, t_{ab}, \nabla_a)$ is a classical spacetime model;*
2. *ϕ and ρ are smooth scalar fields on M (representing the gravitational potential and mass density, respectively);*
3. *$R^a{}_{bcd} = 0$ (spacetime is flat);*
4. *$\nabla^a \nabla_a \phi = 4\pi G \rho$ (4-dim Poisson equation); and*
5. *any time-like curve γ representing the possible trajectory of a test particle must satisfy $\xi^a \nabla_a \xi^b = -\nabla^b \phi$ (equation of motion).*

The idea of an NG-model generalizes to a flat 4-dimensional classical spacetime framework the 3-dimensional Euclidean space-and-time conception of Newtonian gravitation expressed in terms of the standard Poisson equation ($\nabla^2 \phi = 4\pi G \rho$). The distribution of matter (ρ) generates a spacelike gravitational field ($-\nabla^a \phi$), which in turn (via condition 5) deflects all test particles equally from inertial spacetime trajectories. However, unlike the 3-dimensional version, in an NG-model the mathematical object encoding inertial structure—namely, the affine connection, ∇_a—is made explicit, as is its role in the gravitational dynamics.[5]

[4] This definition and that of a GNG-model are adapted from Weatherall (2016a).

[5] It's for this reason that Earman and Friedman (1973) claim that the physical content of Newton's Law of Inertia can only be fully understood and appreciated in the 4-dimensional

2.3 GNG-Models

The transition to GNG is motivated by the recognition that the possible trajectories of accelerated particles picked out by the equation of motion in an NG-model can be equally well characterized as the geodesics in a spacetime equipped with a *curved* affine connection, given appropriate modifications elsewhere. NG and GNG turn out to be observationally equivalent in that, given a distribution of matter ρ, they agree on the possible trajectories of all test particles. That is, in part, what then motivates the question of their theoretical equivalence.

The formal construction of GNG occurs in stages. The mathematical relationship between (a) accelerated trajectories in a classical spacetime with flat affine structure and (b) geodesics in a classical spacetime with curvature is made precise via the following 'geometrization' theorem due to Trautman (1965):

Theorem 1 *Let $(M, h^{ab}, t_{ab}, \nabla_a)$ be a classical spacetime model, ϕ any smooth function on M, and $\overset{\phi}{\nabla}_a = (\nabla_a, C^a{}_{bc})$, where $C^a{}_{bc} = -t_{bc}\nabla^a\phi$.*[6] *Then the following are both true:*

1. *$(M, h^{ab}, t_{ab}, \overset{\phi}{\nabla}_a)$ is a classical space-time model; and*
2. *$\overset{\phi}{\nabla}_a$ is the unique derivative operator on M such that, for all time-like curves γ:*

$$\xi^n \overset{\phi}{\nabla}_n \xi^a = 0 \text{ if and only if } \xi^n \nabla_n \xi^a = -\nabla^a \phi.$$

If we take ϕ to be a gravitational potential, Theorem 1 tells us that the possible motions of particles moving under the influence of a gravitational field in a flat classical spacetime coincide with the motions of particles following

spacetime context.

[6]Two affine connections are always related by a smooth (1,2)-tensor field: given ∇_a and such a tensor field $C^a{}_{bc}$, one can define a new affine connection ∇'_a whose associated Christoffel symbols are given by $\Gamma'^a{}_{\beta\nu} = \Gamma^\alpha{}_{\beta\nu} + C^\alpha{}_{\beta\nu}$. I follow Malament (2012) in expressing this relationship between affine connections by writing $\nabla'_a = (\nabla_a, C^a{}_{bc})$.

geodesics in a classical spacetime equipped with a curved affine connection determined by ϕ (namely, $\overset{\phi}{\nabla}_a$).

What remains to be shown in developing GNG is that there is a way of connecting ρ and $\overset{\phi}{\nabla}_a$ directly via a replacement for the Poisson equation, so that (what we would otherwise call) the gravitational effects of matter are encoded dynamically into the geometry of spacetime. It's a consequence of Theorem 1 that in the special case in which the original spacetime is flat (i.e., ∇_a is such that $R^a{}_{bcd} = 0$), for any non-negative function ρ on M,

$$\nabla_n \nabla^n \phi = 4\pi G\rho \quad \text{if and only if} \quad \overset{\phi}{R}_{bc} = 4\pi G\rho t_{bc},$$

where $\overset{\phi}{R}{}^a{}_{bcd}$ is the Riemann curvature tensor associated with $\overset{\phi}{\nabla}_a$ and $\overset{\phi}{R}_{bd}$ is the Ricci tensor defined by $\overset{\phi}{R}_{bd} = \overset{\phi}{R}{}^a{}_{bad}$.[7] This result shows that the observational content of the Poisson equation can be captured in a classical spacetime model with curvature and no gravitational potential. That feature is what characterizes the models of GNG:

Definition 3 *A **GNG-model** is a structure $(M, h^{ab}, t_{ab}, \overset{\phi}{\nabla}_a, \rho)$, where:*

1. *$(M, h^{ab}, t_{ab}, \overset{\phi}{\nabla}_a)$ is a classical space-time model;*
2. *ρ is a smooth scalar field on M (representing the mass density);*
3. *$\overset{\phi}{R}{}^{ab}{}_{cd} = 0$;*
4. *$\overset{\phi}{R}{}^{[a}{}_{[b}{}^{c]}{}_{d]} = 0$;*
5. *$\overset{\phi}{R}_{bc} = 4\pi G\rho t_{bc}$ (replacement for Poisson's equation); and*

[7] It also follows that

1. $\overset{\phi}{R}{}^{ab}{}_{cd} = 0$ (rotation standard), and
2. $\overset{\phi}{R}{}^{[a}{}_{[b}{}^{c]}{}_{d]} = 0$ (curl-freeness),

but neither of these conditions will play an important role in what follows. They do occur in Definition 3 below, however. See Malament (1986, pp.191–192) and Malament (2012, pp.267–269) for additional discussion of these conditions.

6. *any time-like curve γ representing the possible trajectory of a test particle must satisfy $\xi^a \overset{\phi}{\nabla}_a \xi^b = 0$ (equation of motion).*

Here the matter distribution (ρ) is correlated with spacetime curvature ($\overset{\phi}{\nabla}_a$) via condition 5 and test particles traverse geodesic trajectories. There is no gravitational field deflecting particles from inertial motion, as that part of the dynamical picture has now been encoded directly into the geometry. As the matter distribution changes, so too does the affine connection, and thus the structure of spacetime in GNG is not just curved but dynamical.[8]

3 The Argument for Equivalence

But doubts have been raised regarding the physical interpretations associated with NG and GNG suggested above. More specifically, a growing number of philosophers have questioned the adequacy of an interpretation of NG according to which gravitational fields and flat spacetime are basic parts of the physical ontology. Given the parts of the NG formalism that *are* taken to encode basic physical structure, it's but a short step to the claim that NG is really an ontologically misleading reformulation of GNG.[9]

The *argument for equivalence* (as I will call it) turns on the claim that

[8]For every NG-model there is a *unique* observationally equivalent GNG-model. The converse is not true. Trautman (1965)'s 'recovery' theorem shows that for every GNG-model $(M, h^{ab}, t_{ab}, \overset{\phi}{\nabla}_a, \rho)$ *there exists* an NG-model $(M, h^{ab}, t_{ab}, \nabla_a, \phi, \rho)$ such that, for any timelike curve γ in M: $\xi^n \overset{\phi}{\nabla}_n \xi^a = 0$ if and only if $\xi^n \nabla_n \xi^a = -\nabla^a \phi$. But the NG-model is not unique, a fact that underpins the argument for equivalence outlined in the next section. The geometrization and recovery theorems together establish the observational equivalence of NG and GNG.

[9]This line of reasoning is contained, implicitly or explicitly, in Malament (1986), Malament (1995), Malament (2007), Malament (2012), Norton (1995), and Stachel (2007). Much of the argument presented here—particularly concerning the interpretive implications of the relevant symmetry considerations—is also contained in Knox (2014), Weatherall (2016a), and Weatherall (2016b), although there are ways in which the argument considered in the text diverges from their presentations.

there is an unappreciated symmetry in NG, the import of which ought to change our understanding of ∇_a and $-\nabla^a\phi$. Given any pair (∇_a, ϕ) in an NG-model $(M, h^{ab}, t_{ab}, \nabla_a, \phi, \rho)$, there are indefinitely many other pairs (∇'_a, ϕ') in other NG-models $(M, h^{ab}, t_{ab}, \nabla'_a, \phi', \rho)$ encoding the same particle trajectories in M. That is, there are infinitely many other NG-models equipped with distinct flat inertial structures and gravitational fields, but which are nonetheless observationally equivalent. The relationship between these models is made precise in the following theorem:

Theorem 2 *Let $(M, h^{ab}, t_{ab}, \nabla_a, \phi, \rho)$ be an NG-model; ψ be any smooth scalar function on M such that $\nabla^a\nabla^b\psi = 0$; and $\phi' = \phi + \psi$. Then $(M, h^{ab}, t_{ab}, \nabla'_a, \phi', \rho)$ is an observationally equivalent NG-model with $\nabla'_a = (\nabla_a, -t_{bc}\nabla^a\psi)$.*[10]

These models are associated with the same GNG-model but encode different inertia/gravitation splits. Our inability to discern *the* correct inertia/gravitation split seems built into the structure of the theory, making no difference to the theory's observational content. Let us call the $(\nabla_a, \phi) \to (\nabla'_a, \phi')$ transformation characterized in Theorem 2 the *inertia/gravitation symmetry* of NG.[11]

Given this symmetry, the apparent equivalence of NG and GNG follows from a well-rehearsed line of argument. Whenever possible, mathematical objects and structures exhibiting transformations that make no empirical difference ought to be interpreted as *gauge* quantities—as mathematical artifacts or indications of surplus structure—and not as features of a theory's ontology. This is of course how the 3-dimensional gravitational potential Φ is generally presented in standard textbook treatments of Newtonian gravi-

[10] See Malament (2012, pp.274–277).

[11] More precisely, the symmetry is given by the following pair of transformations for smooth ψ: (1) $\Gamma^{\alpha}_{\beta\nu} \to \Gamma'^{\alpha}_{\beta\nu} = \Gamma^{\alpha}_{\beta\nu} - t_{\beta\nu}\nabla^{\alpha}\psi$ and (2) $\phi \to \phi' = \phi + \psi$.

tation: the transformation $\Phi \rightarrow \Phi' = \Phi + C$, where C is a constant function, makes no difference to the gravitational field encoded ($\nabla\Phi = \nabla\Phi'$). The argument for equivalence merely extends this interpretive consideration to the 4-dimensional context of NG's inertia/gravitation symmetry.[12] The availability of the GNG formalism then provides an obvious route for eliminating this surplus structure. Just introduce a gauge-invariant curved affine connection to represent the 'sum' of the inertial and gravitational pieces (adapting the field equation as necessary) and take *it*, rather than ∇_a and ϕ individually (or ∇_a and $\nabla^a\phi$) to carry genuine physical significance—a modification that *just is* GNG. As Knox (2014, p.878) puts it:

> [T]he gravitational field and flat connection are pieces of surplus structure. However, a function of these, I^a_{bc} [$= \Gamma^a_{bc} + t_{bc}\nabla^a\phi$], ... is gauge invariant and should be counted as physical structure. And this invariant quantity has all the properties of a curved connection, with all the right links to the rest of physics (via the equivalence principle) to count as a piece of spacetime geometry. As a result NG itself is best interpreted as a curved spacetime theory, albeit written in a form that obscures its geometrical structure. NG and GNG thus possess the same geometry, and are not an example of the underdetermination of spacetime geometry.

According to the argument for equivalence, then, we ought to interpret any inertia/gravitation split embedded in NG as a matter of descriptive convention, not as an ontologically real distinction between physical field and background spacetime, and (moreover) we ought to see the NG formulation as a whole as an ontologically misleading reformulation of GNG.

[12] See Knox (2014, pp.866, 869–871). Weatherall (2016a, p.1085) seems to endorse this argument as well, and takes the resulting gauge interpretation to be a "natural understanding" (p.1073) of NG. For what I take to be a 3-dimensional version of this argument, see Stachel (2007). I'm grateful to Michel Janssen for bringing Stachel's paper to my attention.

4 Situating Newtonian Gravitation Theory

Even embracing the interpretive principle that underpins the argument for equivalence, I nevertheless maintain that there is a conceptually and methodologically well-motivated understanding of NG on which the argument fails. I do not deny that the preceding approach to NG is an important tool for some tasks, but the alternative developed in this section turns out to be of substantial philosophical relevance in its own right.

4.1 The Embedded Approach

Let us start with the observation that Newtonian dynamics itself is not so much a theory as it is a theoretical framework. That framework was initially proposed in the *Principia*, where the gravitational force law was also posited, but Newton was famously circumspect regarding what physical forces actually exist.[13] Much of the *Principia* is devoted to developing this general framework, with the gravitational force law only playing a central role in the third and final book. Indeed, in the Preface Newton characterizes his dynamical project as that of developing a rigorous mathematical framework to guide the search for forces, writing that

> our present work sets forth mathematical principles of natural philosophy. For the whole difficulty of philosophy seems to be to discover the forces of nature from the phenomena of motions and then to demonstrate the other phenomena from these forces. It is to these ends that the general propositions in books 1 and 2 are directed, while in book 3 our explanation of the system of the world illustrates these propositions. For in book 3, by means of propositions demonstrated

[13] See Sklar (2013, ch. 6) for a brief discussion of Newton's reluctance to interpret the law of gravitation as reflecting a genuine physical force. One of the central claims in Smith (2012)'s magisterial analysis of the *Principia* is that it made respectable the construction of such 'intermediate level theories', which posit mathematical laws but remain silent regarding their underlying mechanisms (see especially pp. 370, 375).

> mathematically in books 1 and 2, we derive from celestial phenomena the gravitational forces by which bodies tend toward the sun and toward the individual planets. Then the motions of the planets, the comets, the moon, and the sea are deduced from these forces by propositions that are also mathematical. If only we could derive the other phenomena of nature from mechanical principles by the same kind of reasoning! For many things lead me to have a suspicion that all phenomena may depend on certain forces by which the particles of bodies, by causes not yet known, either are impelled toward one another and cohere in regular figures, or are repelled from one another and recede. Since these forces are unknown, philosophers have hitherto made trial of nature in vain. But I hope the principles set down here will shed some light on either this mode of philosophizing or some truer one. (Newton, 1999, pp.382–3)

The *Principia* can thus be understood as doing at least three things: (1) outlining a research program of searching for forces; (2) providing a mathematical framework in which to pursue that dynamical program; and (3) proposing one particular constitutive force law—the gravitational one—as a part of that program.[14] The framework itself is quite general, accommodating many different types of constitutive forces, and Newton's hope is clearly to construct a comprehensive theory of matter in motion in terms of a fundamental set of forces. Underpinning this project is a particular conception of the nature and structure of forces: namely, that their dynamical

[14]The idea that Newton developed a general dynamical framework in which to pursue the search for forces, as opposed to developing a specific physical theory, is taken by Smith (2012, pp.369–371) to be an entirely new way of doing physics—what he identifies as the second way (of nine) in which the *Principia* changed physics. For an extensive discussion of how that program was pursued and developed (and amended) in the century after Newton, see Truesdell (1968) and Hankins (1967). This program reached its pinnacle (and most rigid form) at the beginning of the 19th century in the work of the French Newtonians Pierre-Simon Laplace (in physics) and Claude Louis Berthollet (in chemistry), who sought to explain all phenomena via the postulation of central attractive and repulsive forces acting between (ponderable and imponderable) matter. See Fox (1974) for a description of this program, including an account of its demise. I am certainly not the first to use the programmatic feature of the *Principia* for philosophical ends. Kitcher (1981) invokes it in the service of developing his account of explanatory unification, suggesting that the unifying potential of this program played an important historical role in its 18th century acceptance.

effects *compose*. The trajectory of a body experiencing multiple forces will be the composition of the trajectories associated with each force acting in isolation. If this were not the case—if the postulation of one force affected the *dynamical structure* of how other forces interacted with a body—then the very program of searching for forces within a unifying framework would not make sense.

Interpretive discussions of Newtonian *gravitation* theory generally ignore this background context, preferring instead to consider NG in isolation from the larger mathematical framework.[15] Consider that the equation of motion expressed in NG-models above ($\xi^a \nabla_a \xi^b = -\nabla^b \phi$) makes no mention of Newton's second law ($m\xi^a \nabla_a \xi^b = F^b_{\text{net}}$). The former is only adequate if one restricts application of the latter to just the law of universal gravitation: with no particle 'charges' associated with non-gravitational forces, there's no harm in dispensing with the mass term on both sides of the second law. Let us call this way of approaching the theory the *received view* of NG. It arises by narrowing the Newtonian framework to solely the gravitational force law and then discarding those features not needed to sustain the resulting structure.

However, we need not embrace this parochialism. We might instead choose to situate NG within a larger context, and view the formalism as one part of the larger project outlined in the *Principia*. Other structures exist within the Newtonian framework, positing other constitutive force laws hypothesized to be dominant in other domains; they, too, form part of the

[15] Here I'm using 'Newtonian gravitation theory' (and 'NG') to designate a particular *mathematical* formalism. Throughout this paper I have been reluctantly perpetuating an equivocation rife within the philosophical literature, namely, that between 'Newtonian gravitation theory' (and 'NG') as a bonafide physical theory and as the bare uninterpreted formalism associated with that theory. I trust that context will disambiguate, as deplorable as the equivocation is.

broader dynamical program. Thinking of NG as situated or embedded within such a program involves attending to ways in which the NG formalism is or might be connected to the broader project of searching for forces, and thus to possible relationships NG might bear to Newtonian theories of non-gravitational forces. (Of course, viewing NG through this lens does not pre-judge the question of how the 'gravitational force' itself is to be understood, as to see NG as part of a larger program is not to commit to interpreting all constitutive force laws as representing genuine forces.) Let us call this way of thinking about NG the *embedded view* of the theory. It arises by considering NG as a special application of the more general Newtonian mathematical framework.

4.2 Reinterpreting the Gravitational Formalism

Although viewing NG in this way does not change the class of models associated with the theory's formal structure—the possible worlds of NG remain those characterized formally by NG-models—it does suggest interpretive considerations not present when the NG formalism is viewed through the lens of the received view. Seeing NG as a piece of the broader Newtonian program involves understanding that formalism as one part of a larger formalism incorporating various (perhaps to-be-determined) force laws. Insofar as we expect NG to be meshed with other Newtonian theories, we have good reason to interpret mathematical structures within the NG formalism as physically significant if we expect that they will be taken as such within the larger, more encompassing formalism. For the idea that NG is part of a larger dynamical program amounts *inter alia* to recognizing that NG (however interpreted) is *incomplete* as an account of matter in motion.

That recognition, and insights into how the formalism is likely to be supplemented and potentially revised, ought to inform our understanding of that part of the story we think *is* well represented by NG. To ignore the role certain mathematical structures are likely to play in a more encompassing successor theory is to discard potentially important information relevant to understanding the nature of gravitation itself. We can, if we like, imagine that NG represents a complete account of matter in motion, and screen off interpretive considerations external to the NG formalism; that is precisely what the received view does. But once we decide to consider NG as situated within a larger framework and program, considerations regarding how the NG formalism is to be meshed with other parts of that framework and program become interpretively relevant.

When we take these additional interpretive considerations into account, we find that the inertia/gravitation split carries physical significance we might not have thought it had—that there are conceptual grounds for fixing it one way and not another. Consider the result of meshing NG with an arbitrary Newtonian theory positing a single constitutive force law. Let us call the resulting theory $NG+1$. It's models are defined as follows:

Definition 4 *An* ***(NG+1)-model of Newtonian dynamics*** *is a structure* $(M, h^{ab}, t_{ab}, \nabla_a, \rho, \rho^*, \phi, \phi^*)$*, where:*

1. $(M, h^{ab}, t_{ab}, \nabla_a)$ *is a classical space-time model;*
2. $\rho, \rho^*, \phi, \phi^*$ *are smooth scalar fields on* M*;*
3. $R^a{}_{bcd} = 0$*;*
4. $\nabla^a \nabla_a \phi - 4\pi G\rho = 0$*;*
5. $f^*(\rho^*, \phi^*) = 0$ *(arbitrary second constitutive force law); and*
6. *any time-like curve* γ *representing the possible trajectory of a test particle of mass* m *and charge* q^* *must satisfy* $m\xi^a \nabla_a \xi^b = -m\nabla^b \phi - q^* \nabla^b \phi^*$ *(equation of motion).*

Here ρ^* and ϕ^* are naturally interpreted as representing the charge density and force potential of the non-gravitational force, respectively. The pairwise transformation $(\nabla_a, \phi) \to (\nabla'_a, \phi')$ of Theorem 2 remains a symmetry of $NG+1$, in the sense that any particle trajectory γ satisfying $m\xi^a \nabla_a \xi^b = -m\nabla^b \phi - q^* \nabla^b \phi^*$ will also satisfy $m\xi^a \nabla'_a \xi^b = -m\nabla'^b \phi' - q^* \nabla'^b \phi^*$. However, despite this, there is an important but subtle difference in how these two equations require us to think about the nature and dynamical structure of forces.

Consider a particle initially at rest at the origin of a Cartesian coordinate system ($\Gamma^\alpha_{\beta\nu} = 0$), and which experiences two forces: (1) a constant gravitational force of unit magnitude in the $+x$ direction, $F^a_{(g)} = -m\nabla^a \phi \to m(-\partial^0 \phi, -\partial^1 \phi) = (0, 1)$; and (2) a force coupled to the charge q^*, also constant in the $+x$ direction, with a magnitude twice that of $F^a_{(g)}$, $F^a_{(*)} = -q^* \nabla^a \phi^* \to q^*(-\partial^0 \phi^*, -\partial^1 \phi^*) = (0, 2)$.[16] The equation of motion

$$m\xi^a \nabla_a \xi^b = -m\nabla^b \phi - q^* \nabla^b \phi^*$$

takes the following coordinate-dependent form:

$$\begin{aligned} \tfrac{d^2 x^\alpha}{dt^2} + \Gamma^\alpha_{\beta\nu} \tfrac{dx^\beta}{dt} \tfrac{dx^\nu}{dt} &= -\partial^\alpha \phi - \partial^\alpha \phi^* \\ \tfrac{d^2 x^\alpha}{dt^2} &= -\partial^\alpha \phi - \partial^\alpha \phi^*, \end{aligned}$$

where we use the knowledge that t is an affine parameter. The $\alpha = 0$ ($\alpha = t$) equation is trivial, but the $\alpha = 1$ equation is

$$\frac{d^2 x^1}{dt^2} = -\partial^1 \phi - \partial^1 \phi^* = 1 + 2 = 3,$$

the solution of which is

$$x^1(t) = \frac{3}{2} t^2,$$

[16]For simplicity we'll set $m = q^* = 1$, and ignore two spatial dimensions.

just as one would expect for a particle experiencing a constant net force.

Now suppose we apply an inertia/gravitation transformation $(\nabla_a, \phi) \rightarrow (\nabla'_a, \phi')$ of Theorem 2:

$$\Gamma^{\alpha}_{\beta\nu} \rightarrow \Gamma'^{\alpha}_{\beta\nu} = \Gamma^{\alpha}_{\beta\nu} - t_{\beta\nu}\nabla^{\alpha}\psi$$

$$\phi \rightarrow \phi' = \phi + \psi,$$

where $\nabla^a\nabla^b\psi = 0$. For concreteness, let's say $\psi(t,x) = -2x$, so that $\partial^0\psi = 0$ and $\partial^1\psi = -2$. Within our coordinate system,

$$\Gamma'^{\alpha}_{\beta\nu} = \Gamma^{\alpha}_{\beta\nu} - t_{\beta\nu}\nabla^{\alpha}\psi = -t_{\beta\nu}\nabla^{\alpha}\psi = -t_{\beta\nu}\partial^{\alpha}\psi.$$

Given the degenerate signatures of t_{ab} and h^{ab}, the only non-vanishing component of $\Gamma^{\alpha}_{\beta\nu}$ is $\Gamma^{1}_{00} = -\partial^1\psi = 2$. The transformed equation of motion

$$m\xi^a\nabla'_a\xi^b = -m\nabla'^b\phi' - q^*\nabla'^b\phi^*$$

takes the coordinate-dependent form

$$\begin{aligned} \tfrac{d^2x^\alpha}{dt^2} + \Gamma'^{\alpha}_{\beta\nu}\tfrac{dx^\beta}{dt}\tfrac{dx^\nu}{dt} &= -\partial^\alpha\phi' - \partial^\alpha\phi^* \\ &= -\partial^\alpha\phi - \partial^\alpha\psi - \partial^\alpha\phi^*. \end{aligned}$$

Again the $\alpha = 0$ equation is trivial, but $\alpha = 1$ gives:

$$\begin{aligned} \tfrac{d^2x^1}{dt^2} + \Gamma'^{1}_{00}\tfrac{dx^0}{dt}\tfrac{dx^0}{dt} &= -\partial^1\phi - \partial^1\psi - \partial^1\phi^* \\ \tfrac{d^2x^1}{dt^2} - \partial^1\psi &= -\partial^1\phi - \partial^1\psi - \partial^1\phi^* \\ \tfrac{d^2x^1}{dt^2} - \partial^1\psi &= -\partial^1\psi + 3 \\ \tfrac{d^2x^1}{dt^2} &= 3, \end{aligned}$$

the solution of which is

$$x^1(t) = \frac{3}{2}t^2,$$

just as it was in the untransformed situation. This of course just confirms what was previously stated: the inertia/gravitation symmetry is preserved in $NG+1$.

Yet the transformation has effected a subtle change in how we're required to think about the theory's dynamical structure. Prior to performing the transformation, for the gravitational force acting alone, the equation of motion is simply

$$\frac{d^2x^1_{(g)}}{dt^2} = -\partial^1\phi = 1,$$

the solution of which is

$$x^1_{(g)}(t) = \frac{1}{2}t^2.$$

Similarly, the equation of motion for the second force acting in isolation is

$$\frac{d^2x^1_{(*)}}{dt^2} = -\partial^1\phi^* = 2,$$

the solution of which is

$$x^1_{(*)}(t) = t^2.$$

Evidently, the total trajectory is the sum of these two dynamical effects:

$$x^1(t) = x^1_{(g)}(t) + x^1_{(*)}(t) = \frac{1}{2}t^2 + t^2 = \frac{3}{2}t^2.$$

However, the same cannot be said *after* the inertia/gravitation transformation is performed. The equation of motion for the gravitational force acting alone would again be $\frac{d^2x^1_{(g)}}{dt^2} = 1$, with a solution of $x^1_{(g)}(t) = \frac{1}{2}t^2$. But the equation of motion for the second force is

$$\begin{aligned}
\frac{d^2x^1_{(*)}}{dt^2} + \Gamma'^1_{\beta\nu}\frac{dx^\beta_{(*)}}{dt}\frac{dx^\nu_{(*)}}{dt} &= -\partial^1\phi^* \\
\frac{d^2x^1_{(*)}}{dt^2} - \partial^1\psi &= 2 \\
\frac{d^2x^1_{(*)}}{dt^2} + 2 &= 2 \\
\frac{d^2x^1_{(*)}}{dt^2} &= 0,
\end{aligned}$$

with solution

$$x^1_{(*)}(t) = t.$$

Here we can clearly see that

$$\begin{aligned} x^1(t) &\neq x^1_{(g)}(t) + x^1_{(*)}(t) \\ \tfrac{3}{2}t^2 &\neq \tfrac{1}{2}t^2 + t. \end{aligned}$$

Post-transformation, then, the total dynamical effect of both forces is *not* the sum of the dynamical effects of each force taken independently. In the transformed equation of motion, the structure of how the second force interacts dynamically with the particle depends upon the presence of the gravitational force, and thus the two forces can't be thought to independently compose in the way required by the search-for-forces program. One thus has good theoretical grounds not to see all inertia/gravitation splits as on an equal footing. Indeed, the dynamical structure governing the second force will vary across different implementations of the transformation. Only for one particular split will the effects of forces compose: namely, the split (∇_a, ϕ) corresponding to $\psi = 0$, for which no 'compensation term' is built into the affine connection. This provides the proponent of the embedded approach with a compelling theoretical reason to fix the inertia/gravitation distinction one way and not any other, and thus to interpret one pair (∇_a, ϕ) as physically significant in a way that all other pairs (∇'_a, ϕ') are not, despite the inertia/gravitation symmetry.

What we've found, then, is that several mathematical structures familiar from NG-models play importantly different roles in (NG+1)-models, owing to the addition of a second force. There is no temptation to interpret ∇_a or $-\nabla^b\phi$ as mathematical artifacts, as the arbitrariness of the inertia/gravitation distinction that loomed large in the argument for equivalence does not seem conventional when NG is understood through the lens of the embedded approach. There are good theoretical grounds for fixing the split one unique way. But if there are good grounds for taking ∇_a and $-\nabla^b\phi$ as physically significant within the NG+1 formalism, then, as argued above, there are good grounds *on the embedded approach* for taking them as physically significant within the NG formalism as well. For however NG is supplemented and meshed with other parts of the Newtonian framework, it will involve the postulation of additional force laws in a manner that generates (NG+1)-models. Insofar as NG is viewed on this approach as attributing a flat and fixed affine structure to spacetime, whereas GNG attributes a curved and dynamical structure, they present us with competing and theoretically *inequivalent* accounts of the world. Of course, nothing I have said allows us to identify, in the context of NG-models, exactly which (∇_a, ϕ) split represents the genuine distinction between inertial structure and gravitational field. That should quite rightly raise empiricist hackles in all of us. Those concerns are defeasible, however. I've argued that there's a way of thinking about NG according to which we have *good reason* to take the inertia/gravitation split as physically significant, even though exactly how that split is to be drawn remains epistemically inaccessible in the context of NG alone.

5 Methodology: Barrett on Approximate Truth

Let us now turn to the first of two philosophical implications of this alternative approach.

The example of Newtonian gravitation theory is often used to motivate or illustrate general philosophical theses about the methodology of science, and the difference between the received and embedded views of NG matters in such contexts. As an extended case study, consider Barrett (2008)'s account of approximate truth in terms of descriptive nesting. Barrett is motivated by a puzzle within contemporary physical theorizing: quantum mechanics and special relativity are, when taken together, false, and yet they are each taken to provide us with physical knowledge in virtue of being, in some sense, approximately true. How can we make sense of such an attitude, given that we do not comprehend the exact ways in which each theory gets things right (or wrong)? This question is only made more confounding when one reflects on the fact that different proposals for resolving the underlying tension—that is, for resolving the relativistic quantum measurement problem—locate the sources of error in different places. So if we know that the theories cannot both be right, but have no grip on where each one is going wrong, what meaning can be given to the idea that they are each approximately true? In what sense is each theory providing us with physical knowledge?

Barrett's proposal is that we understand such claims of approximate truth in terms of descriptive nesting, a relationship designated to hold between a theory and its successor. Just as we consider Galileo's (strictly false) claim that "the sun is stationary, at the center of the universe, with the earth revolving around it" (p.215) to be approximately true on account of the fact that modern celestial mechanics contains claims that could be construed as rough translations (or "refinements") of Galileo's, so more generally the claims of a discarded theory are thought to be *descriptively nested* within its (less error-riddled) successor to the extent that the new theory is understood as endorsing rough translations of its predecessor's claims. Nesting obviously comes in degrees: some claims but not others might be descriptively nested within a successor. If the nesting relations are "sufficiently rich", Barrett wants to say that the discarded theory is approximately true relative to its successor.[17]

Applied to the predicament of contemporary physical theorizing, Barrett's idea is the following: to say that quantum mechanics and special relativity are approximately true, even though jointly false in a way that we do not fully understand, is to predict that robust descriptive nesting relations will hold between those theories and their to-be-developed successor. In this way we can make sense of the idea that both theories are approximately true and provide us with physical knowledge, even though we are not yet in a position to say exactly which parts of each theory are making those contributions.

Why think contemporary theories *are* approximately true in the proposed sense? Why expect that sufficiently rich descriptive nesting relations will hold between quantum mechanics, special relativity, and their joint successor?[18] For Barrett the answer is methodological:

> There is a standing explanatory demand on future physical theories that they should characterize the descriptive errors as well as account for the predictive and explanatory successes of our current theories insofar as possible given other desired virtues...[R]elatively rich explanations of this feature of our current theories can be given in the context of descriptive nesting relations characterized by the descriptive features of our current theories preserved in subsequent theories and the senses in which those features are preserved. (p.217)

As an illustration of how nesting relations succeed in discharging this explanatory burden, Barrett cites the relationship between Newtonian gravitation theory and the General Theory of Relativity (GTR). He contends

[17] Barrett's is an unapologetically pragmatic conception of approximate truth, as no claim is made that the (more descriptively accurate) successor theory to which approximate truth is relativized is anywhere close to *the truth itself*.

[18] As Barrett notes (p.217, n.11), there must always be *some* descriptive nesting, if only with respect to various observational claims, otherwise the successor theory would be a complete change of subject.

that central claims of GTR are in fact plausibly interpreted as refinements of important claims within NG. Barrett writes, for example, that "there is a precise sense in which the gravitational field equation of GTR is just the field equation of NG for empty space" (p.219), and argues that NG is plausibly understood as the limit of GTR as relativistic effects are washed out. On this basis he concludes that:

> Each of these descriptive nesting relations provides a precise sense in which one can take a feature of NG to have been preserved in GTR. One might then *in these precise senses* judge the descriptions provided by NG to be approximately true from the perspective of GTR. (p.220)

The nesting relations that ground these approximate truths are then invoked to explain the empirical and explanatory successes of NG, as "the predictions and explanations of GTR converge to those of NG as relativistic effects become negligible" (p.221).

Much of this example and the overall line of argument depends upon how one understands NG. The contrast between received and embedded approaches brings to the fore several subtleties and potentially problematic features of Barrett's view. Most obviously, both aspects of Barrett's example rely crucially on the understanding of GNG and its relationship to NG associated with the received view and the argument for equivalence. In the case of the field equation for empty space, it is the "translation" of the Poisson equation into the GNG formalism that is identical to the Einstein field equation for zero mass density, not the actual Poisson equation central to the NG formalism itself. Similarly, it is GNG and not NG that is, mathematically, the non-relativistic limit of GTR, as it is the geometrized "version" of Poisson's equation that is the limiting form of Einstein's field equation. Neither of these claims establish any nesting relationship at all between NG and GTR unless NG and GNG are taken to be theoretically equivalent (with GNG as the ontologically perspicuous formulation[19]). In the case of the non-relativistic limit, for example, Barrett writes:

> ...[O]ne can take GNG to be the limiting description of the world described by GTR as one gradually eases relativistic constraints. Since geometric descriptions in GNG are translatable into force descriptions in NG, this provides an especially compelling sense in which NG might be said to be approximately true from the perspective of GTR. (p.220)

This passage wouldn't make sense unless the invoked geometry-to-force translations were understood both literally and as revealing the genuine physical content of NG.

Now if the NG-GTR case is intended to illustrate that descriptive nesting occurs in actual scientific practice, as Barrett's discussion seems to suggest, the example is a non-starter. For as a historical matter, the received view of NG was never believed by anyone. Those physicists working on gravitation in the wake of the *Principia* adopted an attitude much closer to the embedded approach: although at times they confined their attention to just the gravitational force law, they kept in view the connection of that law to the search for forces within the broader Newtonian framework. Newton, for whom the gravitational force law is obviously the centerpiece of book 3's discussion of celestial dynamics, freely speculates about non-gravitational forces that might also contribute to planetary motions.[20] And later, in the 18th century, the possibility of introducing non-gravitational forces into celestial dynamics played a central role in ongoing attempts to reconcile the

[19] Even if NG and GNG were taken to be theoretically equivalent, the nesting relationships would break down if NG were taken to perspicuously represent the underlying physical ontology.

[20] See, e.g., Newton (1999, p.880), where he speculates that the precession of the lunar orbit might be due to forces emanating from the Earth's magnetic field. The investigation of the role of non-gravitational forces in the dynamics of celestial phenomena dovetails nicely with Truesdell (1968)'s claim that the interest of Newton and the later Newtonians in resistance forces in continuous media—all of book 2 of the *Principia*!—was strongly motivated by its perceived relevance to celestial mechanics.

law of gravitation with astronomical observation. Indeed, the possibility of introducing additional forces plays a decisive role in Smith (2012)'s account of how evidence works in both the *Principia* and the subsequent gravitational research it inspired. He writes that

> [a] long tradition of carelessly talking about evidence in celestial mechanics as if it were straight-forwardly hypothetico-deductive has obscured the extent to which *the focus of ongoing research has been on questions about further forces.* (p.382; my emphasis)

So even when it came to celestial phenomena, where gravitational effects turn out to dominate all others, physicists who endorsed NG kept in view the possibility of incorporating non-gravitational forces. That is, they adopted something like the embedded view—the view on which NG and GNG are theoretically distinct. They would not have endorsed the force-to-geometry "translations" needed to make sense of NG and GTR as a case of descriptive nesting.[21] The idea that physicists embraced something like the received view is a philosopher's myth.

On the other hand, one might not care about inter-theoretic relations instantiated by theories actually believed in the history of science. Historical accuracy may not be an overriding concern. In treating Galileo's claims as 'refined' by successor theories, Barrett emphasizes that a certain amount of interpretational flexibility must be employed if we are to make sense of those claims as approximately true (pp.215–216). We must be willing to reinterpret them in a way that resonates with modern celestial mechanics, but also invariably diverges somewhat from Galileo's own understanding. Something similar might be said of Newtonian gravitation: to see its central claims as approximately true, we must be willing to reconstrue them in a

[21] Of course, it's certainly true that they didn't have the mathematical and conceptual machinery needed to formulate GNG. But the point is that they also endorsed an approach to NG on which they wouldn't have been considered theoretically equivalent anyway.

way that is at variance with the meanings historically attached to them. On this reply the proposed descriptive nesting between NG and GTR was never intended to capture a strict historical relationship between theories.

But the degree of interpretational flexibility needed to make sense of NG (or central parts) as approximately true is much more extreme than that needed to understand Galileo's claims. It is not that central NG claims must be construed slightly differently from their ordinary (or historical) meanings. They must be so completely re-understood as to be unrecognizable as construals of claims based on the NG formalism itself. (This is, of course, exactly what the proponent of the equivalence argument advocates.) Consider that in order to understand NG claims about particle accelerations arising from gravitational forces as being descriptively nested within GTR, those claims must be re-construed—when put through the force-to-geometry GNG translation scheme—as being about particles that are *not* accelerated and *not* experiencing forces! This degree of interpretational flexibility can hardly serve as an adequate basis for understanding how our best contemporary theories are approximately true. For it is consistent with this reading of Barrett's analysis that the parts of contemporary theories taken by a successor theory to be approximately true might be so radically reconstrued that, were we to be presented with the reconstruals now, we wouldn't even recognize them as *parts* of our current theories. It is difficult to see how an inter-theoretic relationship that countenances such a possibility could plausibly count as an analysis of approximate truth. Although Barrett may be right that we aren't in a position to say exactly which parts of current theories will turn out to be approximately true, in saying that they *are* approximately true we generally assume that those (unknown) parts will at least be recognizable to us *as* parts of current theories. It's unclear what the point would be of calling them *approximately true* otherwise. The approach to NG needed to make good on the claim of descriptive nesting—the received view—thus appears to render Barrett's general account of approximate truth inadequate.

The embedded approach to NG also makes salient a crucial lacuna in Barrett's methodological argument that we ought to expect (important parts of) current physical theories to be descriptively nested within their successors, and thus that we're entitled to view current theories as approximately true. That argument turned on a particular norm governing theory construction: namely, the expectation that future theories should "account for the predictive and explanatory successes of our current theories" (p.217). Barrett's claim was that such a demand naturally leads to rich nesting between theories, for "descriptive nesting relations...can be expected to help explain both the explanatory and predictive successes and failures of older theories relative to newer theories" (p.221). But the embedded view suggests an alternative explanation of empirical success.[22] Without endorsing any equivalence between NG and GNG, we can explain (from the perspective of GTR) the empirical success of NG simply by noting (a) that GNG is the non-relativistic limit of GTR, and (b) that a precise mathematical relationship holds between the models of NG and GNG—given by Trautman (1965)'s 'geometrization' and 'recovery' theorems—to the effect that identical mass distributions give rise to the same particle trajectories in each formalism. More generally, a theory may be successful (from the view-

[22]I am skeptical that there is any norm regarding the *explanatory* successes of a predecessor theory. Even if we expect the oxygen theory to explain the empirical successes of the phlogiston theory, for example, we surely don't expect it to explain its *explanatory* successes. After all, the oxygen theory denies that phlogiston exists, so explanations invoking phlogiston aren't really successes at all. Indeed, Kitcher (1981, p.730) notes that Lavoisier explicitly denied that the phlogiston theory possessed genuine explanatory power.

point of its successor) because it uses a particular mathematical formalism that, while not bearing a limiting relationship to the successor, replicates the empirical predictions of some *third* formalism that *does* bear a limiting relationship to the successor.[23] This provides an explanation of the empirical successes of the predecessor without descriptively nesting any of the predecessor's substantive theoretical claims within the successor theory. So while we might countenance the methodological norm Barrett identifies, and agree that some relationship ought to obtain between current theories and their to-be-developed successor, it need not be a relationship of descriptive nesting. At the very least, the embedded view illustrates how the connection between theory construction and descriptive nesting is a good deal looser than Barrett suggests.

Let us step back from Barrett's account to consider the broader philosophical lesson: the two approaches to Newtonian gravitation are not suited to the same philosophical tasks. If we are interested in whether there exists an interpretation of NG according to which it is the non-relativistic ($c \to \infty$) limit of GTR, or whether there is a more general sense in which NG, as it is employed in contemporary physics for making calculations, is the reduction of GTR, then invoking the received view is unproblematic.[24] On the other hand, if NG is being invoked in the service of advancing a thesis about scientific methodology and it matters whether it was ever believed by anyone in the history of science, then I contend that it's the embedded view that's relevant—the view according to which NG is theoretically inequivalent to GNG. Given the status of NG as a standard go-to theoretical

[23]Nothing essential to this explanation turns on the specific details of how Barrett characterizes the limiting relationship between GNG and GTR. It could be equally well adapted to the more general approach taken in Fletcher (2019).

[24]See Malament (1986) and Fletcher (2019) regarding these respective questions.

example within the philosophy of science, this conclusion bears on a great number of discussions and arguments about scientific methodology.[25]

6 Theoretical Equivalence Reconsidered

The contrasting approaches to NG developed in this paper also shed interesting light on the broader issue of theoretical equivalence, not just its specific application to Newtonian gravitation theory. In general, on what basis do we decide whether two superficially different mathematical formalisms are in fact reformulations of a single underlying theory?

One line of thought, going back to Quine (1975) and Glymour (1971, 1977, 1980), is to settle this question by appealing to various formal relations that hold between the sets of models associated with the respective formulations. The two approaches to NG outlined in this paper offer a clear counter-example to such model-theoretic formal approaches. There is a way of thinking about NG-models on which the inertia/gravitation symmetry figures quite centrally in their interpretation and on which NG and GNG are theoretically equivalent. There is also a way of thinking about NG-models on which the inertia/gravitation symmetry does *not* play a central role in their interpretation and on which NG and GNG are *not* theoretically equivalent. Both of these approaches concern the same mathematical formalism and the same set of models, and thus on each approach all of the same formal relationships obtain between the set of NG-models and the set of GNG-models. Yet on only one way of thinking about NG-models are they

[25]For example, the central philosophical moral of this section provides grounds for criticizing Saatsi (2019)'s version of scientific realism, which draws on Barrett's analysis of the NG-GTR relationship, and also Fletcher (2019)'s concluding discussion of structural realism, which treats GNG as being of relevance to the historical considerations that often motivate that form of realism.

best understood as theoretically equivalent to the set of GNG-models.[26]

Recently, however, formal approaches to theoretical equivalence centered on the mathematical concept of a *category* have come to dominate the philosophical landscape.[27] These approaches are often contrasted with the view developed in Coffey (2014), where theoretical equivalence is taken to be an interpretive matter regarding the physical content associated with different formalisms.[28] Coffey argues not just that formal criteria cannot capture our concept of theoretical equivalence—that formal criteria cannot provide necessary and sufficient conditions for judging theoretical equivalence—but that criteria beyond "interpretive equivalence" are unnecessary.[29] How might this paper's analysis of NG and GNG bear on this dispute?

The criterion of categorical equivalence associates with a formalism not just a set of models, but a set of 'arrows' or mappings between models.

[26]Other sources of skepticism regarding Quine's and Glymour's accounts have been raised in Sklar (1982), Coffey (2014), Barrett and Halvorson (2016), and Weatherall (2016a).

[27]The category-theoretic criterion has its origins in Halvorson (2012) and Weatherall (2016a), but has since been advocated or extended in: Barrett (2015), Barrett (2018), Barrett (2019), Halvorson (2019), Halvorson and Tsementzis (2017), Hudetz (2019), Nguyen *et al.* (2020), Rosenstock *et al.* (2015), Weatherall (2016c), Weatherall (2017), and Weatherall (2019b). See Weatherall (2019b, p.9, n.1) for an explanation of the somewhat confusing chronology of this view.

[28]See also Sklar (1982) and, for a recent extension of the view, Nguyen (2017). As Weatherall (2019a) notes, to the extent that *empirical* equivalence is an interpretive matter, the category-based formal approaches in question are not *purely* formal. One must be able to say that the formalisms pertain to the same domain of phenomena, say, before one's preferred criterion can be invoked. Still, these approaches are formal in the sense that, having agreed on empirical equivalence, the further question of theoretical equivalence is then determined on the basis of whether certain formal relations obtain.

[29]The phrase "interpretive equivalence" is perhaps confusing, as it suggests that two formalisms are theoretically equivalent if and only if they are given the same interpretations. While Coffey seems to take this as true, the phrase 'the same interpretations' is also likely to mislead: as something that mediates between mathematics and ontology, one interpretation might be radically different from another, and it might be *in virtue of that fact* that their respective formalisms are taken to represent the same underlying physical content.

Together, the entire structure forms a type of category.[30] In the context of physical theories, one can think of a category as taking a set of models and adding a specification of which models are representationally equivalent (as given by the arrows). According to the criterion of categorical equivalence, two formalisms are theoretically equivalent if and only if their associated categories are isomorphic *modulo* representational equivalence—that is, if and only if they are isomorphic provided one deliberately fails to distinguish (or ignores the differences) between models that are related by arrows.[31]

The preceding discussion of NG and GNG might be thought to fit quite well with this criterion. The category associated with the received view is one on which models related by the inertia/gravitation symmetry are taken to be representationally equivalent—they are related by mappings in the relevant set of arrows—and that category is equivalent to the one naturally associated with GNG. So by the categorical equivalence criterion, NG and GNG are theoretically equivalent. Similarly, on the embedded view models related by the inertia/gravitation symmetry are not representationally equivalent, and in the associated category those models are *not* related by mappings in the relevant set of arrows. That category is *not* equivalent to

[30] A mathematical *category* is a set of objects and sets of 'arrows' between pairs of those objects that satisfy certain conditions (e.g., associativity of composed arrows, existence of identity arrows for each object). The mathematical details will not be relevant here, but can be found in Weatherall (2016a), Weatherall (2019b), and Halvorson (2012). In the present context, the objects are taken to be models of a formalism and the arrows are various sorts of structure-preserving isometries (or generalizations thereof) between those models.

[31] Such a relationship is weaker than an actual isomorphism and is called an *equivalence*. As one might expect, the criterion of categorical equivalence requires that the equivalence preserves empirical content. Relationships *between* categories—isomorphisms, equivalences, or otherwise—are made mathematically precise or realized via *functors*, which are mappings between categories that take objects to objects and arrows to arrows. Various simple conditions must also be satisfied, but again the details are not relevant and are presented in Weatherall (2016a) and Halvorson (2012).

GNG's, and so NG and GNG are not theoretically equivalent.[32] All of this seems quite simpatico with the view developed in this paper.

Two related observations should give us pause, however. The first is that the construction of categories and the invocation of categorical equivalence doesn't play any role in the arguments for and against theoretical equivalence in sections 3 and 4. That the criterion is satisfied (or not) doesn't add anything to our understanding of those cases, either. The argument for equivalence proceeded, first, by claiming that elements of the NG formalism previously thought to be physically significant in fact ought to be interpreted as gauge quantities, and, second, by positing that an alternative gauge-invariant quantity (e.g., Knox's I^a_{bc}) ought to be interpreted as representing genuine spacetime structure. The judgment of theoretical equivalence then followed simply by noting that the resulting physical picture was precisely the one represented by the GNG formalism, as it is typically interpreted. Arguments regarding how these formalisms ought to be interpreted are what grounded that judgment, which is precisely why the inertia/gravitation symmetry played such a central role.[33] In a similar way, the considerations associated with the embedded view that implied theo-

[32]These claims are developed with much more care and mathematical precision in Weatherall (2016a, pp.1084–1085). The category one might associate with the embedded view is what Weatherall calls $\mathbf{NG}_1$, the objects of which are NG-models and the arrows of which are diffeomorphisms that preserve the classical metrics t_a and h^{ab}, the mass distribution ρ, the derivative operator ∇, and the gravitational potential ϕ. The category of the received view corresponds to what he calls $\mathbf{NG}_2$, the objects of which are NG-models and the arrows of which are given by pairs (χ, ψ), where ψ is a smooth scalar field satisfying $\nabla^a \nabla^b \psi = 0$ and χ is a diffeomorphism that preserves the classical metrics, the mass distribution, the gauge-transformed derivative operator $\nabla' = (\nabla, t_b t_c \nabla^a \psi)$, and the gauge-transformed gravitational potential $\phi' = \phi + \psi$. Weatherall proves that $\mathbf{GNG}$ is categorically equivalent to $\mathbf{NG}_2$ but not $\mathbf{NG}_1$.

[33]None of this is to assert that interpretation is "easy" or that it can be straightforwardly gleaned from a formalism, an attitude Weatherall (2019b) seems to associate with the interpretive approach. Indeed, Knox (2014) is at pains to *argue* that I^a_{bc} ought to be interpreted as representing genuine spacetime structure.

retical *in*equivalence were also inherently interpretive. It was in virtue of the *physical* differences associated with NG and GNG that they were taken to be theoretically inequivalent. Of course, some formal considerations did play an important background role: the 'geometrization' and 'recovery' theorems were needed to establish empirical equivalence, without which the issue of theoretical equivalence would have been moot. But at no point did categorical equivalence play a role in illuminating the NG-GNG relationship.

There is a good reason for this: to even apply the criterion of categorical equivalence, one must first make substantive interpretive judgments. For the criterion can only be applied once specific categories are associated with the different formalisms—that is, once the arrows between models are specified—and that choice reflects a substantive interpretive commitment regarding representational equivalence. What could justify the claim that two models represent the same physical situation, apart from some interpretation regarding the (perhaps partial) physical content represented in each model? The category associated with the received view only makes sense—is only 'natural'—in virtue of how we've chosen to understand the NG formalism. At various points Weatherall seems to acknowledge this fact. When introducing the NG categories of $\mathbf{NG}_1$ and $\mathbf{NG}_2$, he writes that "[s]ince these options correspond to different interpretations of the formalism, I will treat them as prima facie distinct theories" (2016a, p.1085), and elsewhere he notes that, according to the criterion of categorical equivalence, "whether standard and geometrized Newtonian gravitation are equivalent depends on *a prior choice* of whether models of the standard theory associated to a single model of the geometrized theory should be taken as equivalent" (2019b, p.2; my emphasis). Once the background interpretive work needed to set up the NG categories has been done, there isn't any illuminating work for the standard of categorical equivalence left to do.

The second observation is that the reply above on behalf of the proponent of categorical equivalence misjudges the category-theoretic structure most naturally associated with the embedded view of NG. Recall that the embedded view interprets NG in such a way that there is a preferred inertia/gravitation split, despite its empirical inaccessibility. Let us suppose that (∇_a, ϕ) correctly represents the true split. How are we to understand the NG-models that instead invoke other pairs (∇'_a, ϕ'), related to (∇_a, ϕ) by the inertia/gravitation symmetry transformations of Theorem 2? The answer invoked in the initial defense of categorical equivalence is to see them as simply getting the inertial structure of spacetime wrong. This was the basis for associating the embedded view with the category $\mathbf{NG}_1$. There is another and more compelling option, however: to understand NG-models that invoke (∇'_a, ϕ') as representing the *same* inertia/gravitation split as (∇_a, ϕ), only presented in a way that is mathematically misleading. On this view (∇_a, ϕ) and (∇'_a, ϕ') both represent the same inertia/gravitation split, but (∇_a, ϕ) does so in a way that is ontologically or structurally perspicuous—e.g., the geodesics of ∇_a represent genuine inertial trajectories—whereas (∇'_a, ϕ') does so in a way that is representationally *misleading.* Of course, we don't know which is which, but as a conceptual matter that's no barrier to seeing them as representationally equivalent.

This second option is in fact much more congruent with the core physical picture of the embedded view. For on the embedded view, NG has a fixed, flat inertial structure. That is part of the physical content of the theory. Because that structure does not vary from one mass distribution to the next—from one possible physical situation to the next—it would be inappropriate to take models of the theory to permit the modification of that structure, much as it would be inappropriate to allow for models that vary other fixed parts of the physical picture (e.g., the physical relationship encoded in the Poisson equation). Thus, the natural way of understanding NG-models on the embedded view is such that (∇_a, ϕ) and (∇'_a, ϕ') are representationally equivalent, with one a more perspicuous representation of inertial structure than the other. Conceptually, but not epistemically, this is no different in spirit than Knox (2014)'s proposal that (∇_a, ϕ) and (∇'_a, ϕ') both be understood as misleading mathematical representations of a unique curved and dynamical affine structure more perspicuously represented by the gauge-invariant object I^a_{bc}. As Knox sees it, all three should be understood as representationally equivalent.

Now the proponent of categorical equivalence faces a problem. The category most naturally associated with the embedded view turns out to be $\mathbf{NG}_2$, which is precisely the category associated with the received view, according to which the inertia/gravitation split is a gauge quantity without physical significance. So both the received view of NG and the embedded view of NG are associated with the same category of models, and that category is categorically equivalent to the category associated with GNG-models. But only on the received view, and not on the embedded view, is NG theoretically equivalent to GNG. I conclude that categorical equivalence does not provide a sufficient condition for theoretical equivalence.

7 Conclusion

There are two contrasting ways of thinking about NG. The received view arises by discarding all parts of the Newtonian framework not needed for the dynamics of the gravitational force alone. The embedded view arises by considering NG as a special case in the broader dynamical program of searching for forces; thus, I've argued, it retains the interpretive commitments needed to make sense of that program. Only in the former case is the inertia/gravitation symmetry plausibly interpreted as a gauge freedom, and so only in the former case is the argument for theoretical equivalence with GNG applicable. On the embedded view, NG bears a strikingly different relationship to GNG than has been assumed in many recent discussions of Newtonian gravitation.

I make no claim that one of these approaches is somehow correct and the other not. The received view is ubiquitous in foundational discussions, and for good reason. However, the embedded view developed in this paper deserves a seat at the proverbial table. To see NG as part of a larger dynamical program and framework is certainly a coherent and methodologically justifiable way of investigating the structure and content of Newtonian gravitation. In addition to arguing that it makes salient an interpretive consideration not often recognized in foundational discussions, I have argued that reflection on these contrasting approaches (a) provides a new way of thinking about the role and appropriateness of NG as a stock example invoked in discussions of the methodology of science, and (b) suggests a compelling reason to reject categorical equivalence in favor of the interpretive approach to theoretical equivalence.

References

Barrett, Jeffrey. 2008. "Approximate Truth and Descriptive Nesting". *Erkenntnis* **68** (2): 213–24.

Barrett, Thomas. 2015. "On the Structure of Classical Mechanics". *British Journal for the Philosophy of Science* **66**: 801–828.

Barrett, Thomas. 2018. "What Do Symmetries Tell Us about Structure?" *Philosophy of Science* **85**: 617–639.

Barrett, Thomas. 2019. "Equivalent and Inequivalent Formulations of Classical Mechanics". *British Journal for the Philosophy of Science* **70**: 1167–1199.

Barrett, Thomas and Hans Halvorson. 2016. "Glymour and Quine on Theoretical Equivalence". *Journal of Philosophical Logic* **45**: 467–483.

Coffey, Kevin. 2014. "Theoretical Equivalence as Interpretive Equivalence". *British Journal for the Philosophy of Science* **65** (4): 821–844.

Colodny, Robert, ed. 1986. *From Quarks to Quasars*. Pittsburgh: University of Pittsburgh Press.

Earman, John and Michael Friedman. 1973. "The Meaning and Status of Newton's Law of Inertia and the Nature of Gravitational Forces". *Philosophy of Science* **40** (3): 329–359.

Fletcher, Sam. 2019. "On the Reduction of General Relativity to Newtonian Gravitation". *Studies in History and Philosophy of Modern Physics* **68**: 1–15.

Fox, Robert. 1974. "The Rise and Fall of Laplacian Physics". In *Historical Studies in the Physical Sciences*, vol. 4, edited by Russell McCormmach. Princeton, NJ: Princeton University Press, pp. 89–136.

Glymour, Clark. 1971. "Theoretical Realism and Theoretical Equivalence". In *Boston Studies in the Philosophy of Science*, vol. 8, edited by R. Buck. Dordrecht: Springer, pp. 275–88.

Glymour, Clark. 1977. "The Epistemology of Geometry". *Noûs* **11** (3): 227–51.

Glymour, Clark. 1980. *Theory and Evidence*. Princeton, NJ: Princeton University Press.

HALVORSON, HANS. 2012. "What Scientific Theories Could Not Be". *Philosophy of Science* **79**: 183–206.

HALVORSON, HANS. 2019. "Scientific Theories". In *The Oxford Handbook of Philosophy of Science*, edited by Paul Humphreys. New York: Oxford University Press, pp. 585–608.

HALVORSON, HANS and DIMITRIOS TSEMENTZIS. 2017. "Categories of Scientific Theories". In *Categories for the Working Philosopher*, edited by Elaine Landry. New York: Oxford University Press, pp. 402–429.

HANKINS, THOMAS. 1967. "The Reception of Newton's Second Law of Motion in the Eighteenth Century". *Archives Internationales d'Histoire des Sciences* **20**: 42–65.

HAVAS, PETER. 1964. "Four–Dimensional Formulations of Newtonian Mechanics and Their Relation to the Special and General Theory of Relativity". *Reviews of Modern Physics* **36** (4): 938–965.

HUDETZ, LAURENZ. 2019. "Definable Categorical Equivalence". *Philosophy of Science* **86**: 47–75.

KITCHER, PHILIP. 1981. "Explanatory Unification". *Philosophy of Science* **48**: 507–531.

KNOX, ELEANOR. 2014. "Newtonian Spacetime Structure in Light of the Equivalence Principle". *British Journal for the Philosophy of Science* **65**: 863–880.

MALAMENT, DAVID. 1986. "Newtonian Gravity, Limits, and the Geometry of Space". In Colodny (1986, pp.181–201).

MALAMENT, DAVID. 1995. "Is Newtonian Cosmology Really Inconsistent?" *Philosophy of Science* **62**: 489–510.

MALAMENT, DAVID. 2007. "Classical Relavitiy Theory". In *Handbook of the Philosophy of Physics*, edited by Jeremy Butterfield and John Earman. Amsterdam: Elsevier, pp. 229–73.

MALAMENT, DAVID. 2012. *Topics in the Foundations of General Relativity and Newtonian Gravitation Theory.* Chicago, IL: University of Chicago Press.

MARION, JERRY and STEPHEN THORNTON. 1995. *Classical Dynamics of Particles and Systems.* 4th ed. Saunders College Publishing.

Newton, Isaac. 1999. *The Principia: Mathematical Principles of Natural Philosophy*. Berkeley and Los Angeles: University of California Press. Translated by I. Bernard Cohen and Anne Whitman. Orig. 1687.

Nguyen, James. 2017. "Scientific Representation and Theoretical Equivalence". *Philosophy of Science* **84**: 982–995.

Nguyen, James, Nicholas Teh, and Laura Wells. 2020. "Why Surplus Structure Is Not Superfluous". *British Journal for the Philosophy of Science* **71**: 665–695.

Norton, John. 1995. "The Force of Newtonian Cosmology: Acceleration Is Relative". *Philosophy of Science* **62**: 511–522.

Quine, W.V.O. 1975. "On Empirically Equivalent Systems of the World". *Erkenntnis* **9**: 313–28.

Rosenstock, Sarita, Thomas Barrett, and James Owen Weatherall. 2015. "On Einstein Algebras and Relativistic Spacetimes". *Studies in History and Philosophy of Modern Physics* **52**: 309–316.

Saatsi, Juha. 2019. "What Is Theoretical Progress of Science?" *Synthese* **196**: 611–631.

Sklar, Lawrence. 1982. "Saving the Noumena". *Philosophical Topics* **13** (1): 89–110. Reprinted in (Sklar (1985), pp.49–72).

Sklar, Lawrence. 1985. *Philosophy and Spacetime Physics*. Berkeley and Los Angeles: University of California Press.

Sklar, Lawrence. 2013. *Philosophy and the Foundations of Dynamics*. New York: Cambridge University Press.

Smith, George E. 2012. "How Newton's *Principia* Changed Physics". In *Interpreting Newton: Critical Essays*, edited by Andrew Janiak and Eric Schliesser. Cambridge: Cambridge University Press, pp. 360–395.

Stachel, John. 2007. "The Story of Newstein or: Is Gravity Just Another Pretty Force?" In *The Genesis of General Relativity*, edited by Michel Janssen, Jürgen Renn, Tilman Sauer, and John Stachel. Dordrecht: Springer, pp. 1962–2000.

Taylor, John. 2005. *Classical Mechanics*. Herndon, VA: University Science Books.

Trautman, Andre. 1965. "Foundations and Current Problems of General Relativity". In *Lectures on General Relativity*, edited by S. Deser and K.W. Ford. Englewood Cliffs, NJ: Prentice Hall, pp. 1–248.

Truesdell, Clifford. 1968. *Essays in the History of Mechanics.* New York: Springer.

Weatherall, James Owen. 2016a. "Are Newtonian Gravitation and Geometrized Newtonian Gravitation Theoretically Equivalent?" *Erkenntnis* **81** (5): 1073–1091.

Weatherall, James Owen. 2016b. "Maxwell-Huygens, Newton-Cartan, and Saunders-Knox Space-Times". *Philosophy of Science* **83** (1): 82–92.

Weatherall, James Owen. 2016c. "Understanding Gauge". *Philosophy of Science* **83**: 1039–1049.

Weatherall, James Owen. 2017. "Category Theory and the Foundations of Classical Space-Time Theories". In *Categories for the Working Philosopher*, edited by Elaine Landry. New York: Oxford University Press, pp. 329–348.

Weatherall, James Owen. 2019a. "Theoretical Equivalence in Physics, Part 1". *Philosophy Compass* **14**: 1–11.

Weatherall, James Owen. 2019b. "Theoretical Equivalence in Physics, Part 2". *Philosophy Compass* **14**: 1–12.