Are Newtonian gravitation and geometrized Newtonian gravitation theoretically equivalent?

James Owen Weatherall

Department of Logic and Philosophy of Science
3151 Social Science Plaza A, University of California, Irvine, CA 92697

Abstract

I argue that a criterion of theoretical equivalence due to Clark Glymour [*Noûs* **11**(3), 227 (1977)] does not capture an important sense in which two theories may be synonymous. I then motivate and state an alternative condition that does capture the sense of equivalence I have in mind. The principal argument of the paper is that relative to this second condition, the answer to the question posed in the title is yes, at least on one natural understanding of Newtonian gravitation. I conclude with a discussion of some core topics in philosophy of space and time.

Keywords: Theoretical equivalence, Categorical equivalence, Gauge theory, Geometrized Newtonian gravitation

1. Introduction

Are Newtonian gravitation and geometrized Newtonian gravitation (Newton-Cartan Theory) theoretically equivalent?[1] Clark Glymour (1970, 1977, 1980) has articulated a natural criterion of theoretical equivalence and argued that, by this criterion, the answer is no. I intend to show that the situation is more subtle than Glymour suggests, by displaying an important sense of equivalence between theories that Glymour's criterion is not flexible

Email address: weatherj@uci.edu (James Owen Weatherall)

[1] One might immediately worry about what is intended here by "theoretical equivalence". The notion of theoretical equivalence that I am trying to develop in this paper is one of synonymy between theories. In short, two theories should be understood to be equivalent just in case they say the same things about the world, or alternatively, if they ascribe the same structure to the world. There are surely other notions of what it might mean for two theories to be equivalent, and I do not intend to say that the criteria I consider here are the only options, or even the best options for other contexts in which one might be interested in notions of equivalence between theories. But I do want to argue that, for the purposes at hand, the criteria I discuss are fruitful.

enough to capture.[2] My principal argument will then be that, while there are certainly senses in which Newtonian gravitation is not theoretically equivalent to the geometrized theory, there is nonetheless a robust notion of equivalence on which Newtonian gravitation (properly understood) and geometrized Newtonian gravitation *are* equivalent.[3,4,5]

One consequence of these results is that there exists a pair of theoretically equivalent (at least by one standard), realistic theories that nonetheless disagree with regard to features that one might have taken to be essential parts of a spacetime theory. In particular, one might have taken the flatness of spacetime to be an important feature of Newtonian spacetime physics—even a metaphysical commitment of the theory. But the flatness of spacetime is not preserved under the theoretical equivalence relation I describe here. I take it that this observation provides new insight on old debates concerning the epistemology of geometry, and on more general questions concerning the relationship between physical theories and metaphysics.[6]

I will begin by reviewing the two formulations of Newtonian gravitation, with a focus on the results that relate them to one another. I will then move on to describe Glymour's criterion for theoretical equivalence, according to which the two formulations of Newtonian

[2] I should say immediately that I consider the arguments of the present paper to be a friendly amendment to Glymour's work on this topic. The point is that there is a subtlety that may be overlooked when working in the terms Glymour uses, and that if one thinks about theoretical equivalence in slightly different terms, that subtlety becomes manifest.

[3] David Zaret (1980) has also replied to Glymour on this question. But his argument is markedly different than the one presented here, and Spirtes and Glymour (1982) offer what I take to be an effective reply.

[4] As will become clear, a crucial part of the present argument bears a close connection to arguments presented by John Norton (1992, 1995) and David Malament (1995), regarding whether one should understand Newtonian gravitation as a "gauge theory", and what that means for the relationship between Newtonian gravitation and its geometrized reformulation.

[5] A recent paper by Eleanor Knox (2011) addresses topics closely related to those discussed in the present paper. Knox does not address whether the two formulations of Newtonian gravitation are equivalent, but she does argue that geometrized Newtonian gravitation is more perspicuous with regard to the ontological commitments of both theories. There may be a point of disagreement between us, insofar as I will argue here that since there is a sense in which geometrized Newtonian gravitation and standard Newtonian gravitation are synonymous, their apparently different ontological commitments turn out not to reflect meaningful physical distinctions at all, at least in the context of the theories' background theoretical assumptions.

[6] It also bears on recent debates concerning scientific realism and instrumentalism, though I will defer discussion of this connection to future work.

gravitation fail to be equivalent. Next, I will consider Glymour's condition with regard to two formulations of electromagnetism that, I will argue, should be understood to be synonymous. It will turn out that these theories fail to be equivalent by Glymour's criterion of equivalence, which I take to show that there is an important sense in which two theories may be equivalent that Glymour's condition does not capture. In the following section, I will use some basic ideas from category theory to articulate a precise alternative condition that does capture the sense in which the two formulations of electromagnetism are equivalent.[7] I will then return to the question of principal interest in the present paper, arguing that there are two ways of construing standard (nongeometrized) Newtonian gravitation. I will state and prove a simple proposition to the effect that, by the modified criterion, on one of the two ways of representing standard Newtonian gravitation (but not the other), it is theoretically equivalent to geometrized Newtonian gravitation.[8] I will conclude with some brief remarks relating this work to certain core questions in the philosophy of space and time.

2. Two formulations of Newtonian gravitation

The two theories with which I am principally concerned are Newtonian gravitation (NG), expressed in a four-dimensional, covariant formalism, and a variant of Newtonian gravitation due to Élie Cartan (1923, 1924) and Kurt Friedrichs (1927), called "Newton-Cartan theory" or "geometrized Newtonian gravitation" (GNG).[9] In NG, gravitation is a force exerted by massive bodies on other massive bodies. It is mediated by a gravitational potential, and in the presence of a (non-constant) gravitational potential, massive bodies will accelerate. In GNG, meanwhile, gravitation is "geometrized" in much the same way as in general relativity:

[7] Category theory may not be familiar to all readers. For this reason, I include a short appendix with the basic definitions I will rely on in this section.

[8] I suspect this result could be recovered in Glymour's own (non-category theoretic) terms, though the distinction between the two criteria I will discuss is more perspicuous in the category theoretic terms I will use.

[9] The discussion here is necessarily brief. For a systematic discussion of geometrized Newtonian gravitation, see Malament (2012) or Trautman (1965).

the geometric properties of spacetime depend on the distribution of matter in spacetime, and conversely, gravitational effects are manifestations of the resulting geometry. On their faces, the theories appear quite different from one another, though there is a precise sense, which I will state below, in which they are empirically equivalent. There is also a precise sense in which GNG, rather than NG, is the "classical limit" of general relativity, though I will not discuss that topic here.[10]

Despite their differences, the theories share a common geometrical structure. In both cases, spacetime is represented by a manifold, which I will assume throughout is simply connected.[11] This manifold is equipped with two (degenerate) metrics: a temporal metric t_{ab} that assigns temporal lengths to vectors, and a spatial metric h^{ab} that assigns spatial lengths to co-vectors (and, in an indirect way, to vectors whose temporal length vanishes).[12] These are stipulated to be orthogonal, in the sense that $h^{ab}t_{bc} = \mathbf{0}$ everywhere. The signature of the temporal metric guarantees that there exists (at least locally) a covector field t_a such that $t_{ab} = t_a t_b$; in cases where the spacetime is "temporally orientable", this field can be defined globally. In what follows, I will assume that the spacetimes under consideration are temporally orientable and will work exclusively with t_a rather than t_{ab}. The degenerate metric structure of Newtonian gravitation does not uniquely determine a derivative operator, and so one needs to independently identify a derivative operator on the manifold. This derivative operator ∇ is required to be compatible with both metrics, in the sense that $\nabla_a t_{bc} = \mathbf{0}$ and $\nabla_a h^{bc} = \mathbf{0}$ everywhere. These four elements together define a *classical spacetime.*

Definition 2.1. *A* classical spacetime *is an ordered quadruple,* (M, t_a, h^{ab}, ∇)*, where* M *is a manifold,* t_a *and* h^{ab} *are, respectively, mutually compatible temporal and spatial metrics,*

[10] For more on this, see Ehlers (1981), Künzle (1976), and Malament (1986b,a). For a discussion of how one moves from general relativity to NG by way of GNG, and of how certain classical concepts like "gravitational mass" arise in that transition, see Weatherall (2011).

[11] This assumption is adopted for simplicity; an alternative approach would be to always work locally.

[12] Throughout I will use the abstract index notation, which is explained in Penrose and Rindler (1984) and Malament (2012).

and ∇ is a derivative operator compatible with both metrics.

In both theories, the matter content of spacetime is represented by a smooth, symmetric field T^{ab}, called the *mass-momentum field.* This field can be used to define a *mass density* on spacetime by $\rho = T^{ab}t_a t_b$. In both theories, massive point particles can be represented by their worldlines—smooth, curves whose tangent vector fields have non-vanishing temporal length, as determined by t_a. Such fields are called *timelike*; vector fields with vanishing temporal length are called *spacelike.*

In this context, NG can be understood as a theory whose models are classical spacetimes with flat ($R^a{}_{bcd} = \mathbf{0}$) derivative operators,[13] endowed with a scalar gravitational field φ satisfying Poisson's equation, $\nabla^a \nabla_a \varphi = 4\pi\rho$. In the presence of a gravitational field, a massive point particle whose worldline has tangent field ξ^a will accelerate according to $\xi^n \nabla_n \xi^a = -\nabla^a \varphi$. In the geometrized version of the theory, meanwhile, the derivative operator is permitted to be curved and the gravitational field is omitted. The curvature field associated with the derivative operator satisfies a geometrized version of Poisson's equation, $R_{ab} = 4\pi\rho t_a t_b$, and in the absence of any external (i.e., non-gravitational) interactions, massive particles traverse timelike geodesics of this curved derivative operator. These two changes provide the precise sense in which gravity is "geometrized" in GNG. In both cases, the "empirical content" or the "predictions" of the theory can be understood to consist in the trajectories of massive bodies, given a particular mass density on spacetime. On NG, these trajectories correspond to the class of curves whose acceleration at a point is given by $-\nabla^a \varphi$; on GNG, these trajectories correspond to the timelike geodesics of the derivative operator. The two theories are empirically equivalent in the sense that, given a mass density on spacetime, both theories agree on the possible trajectories of bodies.[14]

[13] The Riemann curvature tensor $R^a{}_{bcd}$ is defined as in standard differential geometry. Given a derivative operator, the curvature tensor is the unique tensor such that for any vector field ξ^a, $R^a{}_{bcd}\xi^b = -2\nabla_{[c}\nabla_{d]}\xi^a$.

[14] There is an important caveat here. Given only a mass density on spacetime, *neither* formulation of Newtonian gravitation uniquely determines the trajectories of massive bodies. In both cases, there is a degree

Models of NG and GNG are systematically related. First, it is always possible to "geometrize" the gravitational field on a flat classical spacetime, in the sense that given a model of NG, one can always produce a (unique) model of GNG that agrees on (a) the mass density on spacetime ρ and (b) the trajectories of massive bodies. Since the metrical structure of a classical spacetime is fixed, the models of both theories also agree on that. This translation is given via a result due to Andrzej Trautman (1965).

Proposition 2.2 (Trautman Geometrization Lemma). *Let* $(M, t_a, h^{ab}, \overset{f}{\nabla})$ *be a flat classical spacetime. Let* φ *and* ρ *be smooth scalar fields on* M *satisfying Poisson's equation,* $\overset{f}{\nabla}_a \overset{f}{\nabla}{}^a \varphi = 4\pi\rho$. *Finally, let* $\overset{g}{\nabla} = (\overset{f}{\nabla}, C^a{}_{bc})$, *with* $C^a{}_{bc} = -t_b t_c \overset{f}{\nabla}{}^a \varphi$.[15] *Then* $(M, t_a, h^{ab}, \overset{g}{\nabla})$ *is a classical spacetime;* $\overset{g}{\nabla}$ *is the unique derivative operator on* M *such that given any timelike curve with tangent vector field* ξ^a,

$$\xi^n \overset{g}{\nabla}_n \xi^a = \mathbf{0} \Leftrightarrow \xi^n \overset{f}{\nabla}_n \xi^a = -\overset{f}{\nabla}{}^a \varphi; \qquad \text{(G)}$$

and the Riemann curvature tensor relative to $\overset{g}{\nabla}$, $\overset{g}{R}{}^a{}_{bcd}$, *satisfies*

$$\overset{g}{R}_{ab} = 4\pi\rho t_a t_b \qquad \text{(CC1)}$$

$$\overset{g}{R}{}^a{}_b{}^c{}_d = \overset{g}{R}{}^c{}_d{}^a{}_b \qquad \text{(CC2)}$$

$$\overset{g}{R}{}^{ab}{}_{cd} = \mathbf{0}. \qquad \text{(CC3)}$$

It is also possible to go in the other direction. That is, given a model of GNG, it is possible to recover a model of NG with the same mass density on spacetime, that once again agrees on the trajectories of massive bodies.

Proposition 2.3 (Trautman Recovery Theorem). *Let* $(M, t_a, h^{ab}, \overset{g}{\nabla})$ *be a classical spacetime that satisfies eqs.* (CC1)-(CC3) *for some smooth scalar field* ρ. *Then there exists a*

of freedom corresponding to homogeneous solutions to Poisson's equation. However, a unique collection of trajectories is picked out if one imposes additional boundary conditions. In any case, the claim of empirical equivalence is not affected, since precisely the same underdetermination of the trajectories occurs in *both* cases.

[15] The notation $\nabla' = (\nabla, C^a{}_{bc})$ is explained in Malament (2012, Prop. 1.7.3).

smooth scalar field φ *and a flat derivative operator on* M, $\overset{f}{\nabla}$, *such that* $(M, t_a, h^{ab}, \overset{f}{\nabla})$ *is a classical spacetime;* (G) *holds of any timelike curve; and* φ *and* $\overset{f}{\nabla}$ *together satisfy Poisson's equation,* $\overset{f}{\nabla}_a \overset{f}{\nabla}{}^a \varphi = 4\pi\rho$.

It is crucially important for the following discussion, however, that the pair $(\overset{f}{\nabla}, \varphi)$ is not unique. A second pair $(\overset{f}{\nabla}', \varphi')$ will satisfy the same conditions provided that (1) $\nabla^a \nabla^b(\varphi' - \varphi) = \mathbf{0}$ and (2) $\overset{f}{\nabla}' = (\overset{f}{\nabla}, C^a{}_{bc})$, with $C^a{}_{bc} = t_b t_c \nabla^a(\varphi' - \varphi)$. Note, too, that the recovery theorem holds only if the curvature conditions (CC1)–(CC3) are met. Poisson's equation, condition (CC1), has already been assumed to hold of models of GNG; for present purposes, I will also assume that conditions (CC2) and (CC3) hold in all models of GNG.[16] It worth pointing out that these two assumptions hold automatically whenever we begin with NG and translate to GNG.

3. Glymour on theoretical equivalence

Glymour works out his account of theoretical equivalence in several places (Glymour, 1970, 1977, 1980), in the service of a more general argument that there exist cases where two theories can be empirically equivalent, and yet nonetheless inequivalent in a stronger sense. The notion of equivalence he develops for this purpose is supposed to capture what it might mean for two theories to be "synonymous", in the sense of saying the same things about the world. The underlying intuition is that two theories are synonymous if (1) they are

[16] In principle, one can consider models of GNG that do not satisfy condition (CC3). This leads to a generalized version of geometrized Newtonian gravitation. See Ehlers (1981), Künzle (1976), and Malament (2012). If one *does* relax this condition, NG and GNG are not synonymous by any of the criteria I consider here.

empirically equivalent,[17] and (2) they are mutually inter-translatable.[18] There is good reason to think that this first requirement is often a slippery one. What is meant to be the empirical content of a theory may be open to interpretation, or dependent on the way in which predictions and the associated data are represented.[19] That said, at least in some cases, it *is* possible to say what the empirical content of a theory is supposed to be, and moreover to determine that two theories are empirically equivalent. In the case of NG and GNG, for instance, there is a precise sense in which the two theories are empirically equivalent, insofar as they agree on the trajectories of bodies in the presence of a background matter distribution.

The second requirement, of mutual inter-translatability between theories, can be made precise more generally, at least for a first-order theory. Suppose that L and L^+ are first-order signatures (a signature is just the set of non-logical symbols of a language), with $L \subseteq L^+$. An explicit definition of a symbol in L^+ in terms of L can be understood as a sentence in L^+ that asserts the equivalence between that symbol (appropriately used) and some formula in L. To take an example, for any $n-$ary relation symbol R in L^+, and any list of n variables $\bar{x}$, an *explicit definition of* R *in* L is a sentence of the form

$$\forall \bar{x}(R\bar{x} \leftrightarrow \varphi(\bar{x})),$$

where φ is a formula of L with at most n free variables. One can similarly define the forms

[17] Glymour does not explicitly state that empirical equivalence is a necessary condition of theoretical equivalence, though all of the cases he considers *are* empirically equivalent. Nonetheless, one can imagine two theories that are mutually inter-translatable, but which are interpreted in such a way that the predictions of the first theory are translated into parts of the second theory that are not predictions, and vice-versa, so that the two theories should be understood to be empirically inequivalent. It certainly seems that such pairs of theories, if they exist, should not be considered theoretically equivalent. Sklar (1982) also makes the point that empirical equivalence is a substantive additional constraint.

[18] There is a sense in which "mutually inter-translatable" is ambiguous. See bellow, especially footnote 24.

[19] See, for instance, van Fraassen (2008).

of explicit definitions of constants and functions of L^+ in terms of L.[20] Suppose, then, that one has a theory T in L. By appending explicit definitions of (all of the) symbols in L^+/L to T, we can extend T to a theory in L^+. The resulting theory is a *definitional extension of T in L^+*.[21]

The definitional extensions of a theory can be used to define a precise notion of when two first-order theories are mutually inter-translatable.

Definition 3.1. *Suppose T_1 and T_2 are first-order theories in signatures L_1 and L_2, respectively, with $L_1 \cap L_2 = \emptyset$.*[22] *Then T_1 and T_2 are* definitionally equivalent *if and only if there are first order theories T_1^+ and T_2^+ in $L_1 \cup L_2$ such that T_1^+ is a definitional extension of T_1, T_2^+ is a definitional extension of T_2, and T_1^+ and T_2^+ are logically equivalent.*

Definitional equivalence does track an intuitive notion of inter-translatability, since given any pair of definitionally equivalent theories T_1 and T_2 and a formula φ in the language of T_1, it is always possible to translate φ into a formula in the language of T_2, and then back into a formula in the language of T_1 that is T_1-provably equivalent to φ.

Definitional equivalence is the standard of equivalence that Glymour proposes as a standard of synonymy for two (empirically equivalent) first-order theories.[23] In the case of physical theories, however, we typically do not have first-order formulations available to work with, and so the syntactic characterization of definitional equivalence given in definition 3.1 is impractical. But in general, we do know how to work with the models of a physical theory. So instead of adopting definitional equivalence directly, Glymour expresses

[20] Explicit definitions of functions and constants entail additional sentences in L. These entailments are called the "admissibility conditions" of an explicit definition. See Hodges (1993, Ch. 2.6) and fn. 21.

[21] Actually, simply adding explicit definitions of all of the symbols is not quite enough. One also has to require that for any constant or function symbol S in L^+/L, $T \vdash \chi$, where χ is the admissibility condition for S. See Hodges (1993, Ch. 2.6) and fn. 20.

[22] See Tarski and Givant (1987). If L_1 and L_2 are allowed to have non-empty intersection, definitional equivalence fails to be an equivalence relation, because it is not transitive. To see this, consider the three theories $T_1 = \{\forall x, y(xRy \leftrightarrow x = y)$, $T_2 = \{\forall x, y(xRy \leftrightarrow \neg(x = y))\}$, $T_3 = \{\forall x, y(xR'y \leftrightarrow x = y)\}$. But this is not a substantive problem: if the theories have overlapping signatures, one can always generate a new theory by simply modifying the symbols in one of the signatures.

[23] Glymour is hardly alone in adopting definitional equivalence as a standard of synonymy for first-order theories. For instance, see the classic work by de Bouvere (1965b,a). I attribute the proposal to Glymour, however, because he extends this first-order definition to physical theories.

his condition in terms of a model-theoretic consequence of definitional equivalence, which can be stated as follows: Suppose T_1 and T_2 are definitionally equivalent theories in signatures L_1 and L_2 respectively, and suppose that A_1 is an L_1-structure that forms a model of T_1. Then it is always possible to expand A_1 into an $L_1 \cup L_2-$structure A that forms a model of T_1^+, the definitional extension of T_1 that realizes the equivalence. Since T_1^+ and T_2^+ (the extension of T_2) are logically equivalent, A is also a model of T_2^+. We can thus turn A into a model A_2 of T_2 by simply restricting A to symbols in L_2. The whole process can then be reversed to recover A_1. In this sense, definitionally equivalent theories "have the same models" insofar as a model of one theory can be systematically transformed into a model of the other theory, and vice versa. Note that it is essential for this characterization that one can go from a model A_1 of T_1 to a model A_2 of T_2, and then back to the *same* model A_1 of T_1.[24]

Using this model-theoretic characterization, and limiting attention to terms natural to a covariantly expressed field theory, Glymour's condition for theoretical equivalence can thus be stated as follows.[25]

Condition 1. *Two theories T_1 and T_2 are theoretically equivalent just in case for every model M_1 in T_1, there exists a unique model M_2 in T_2 such that (1) M_1 and M_2 are empirically equivalent, and (2) the geometrical objects associated with M_2 are uniquely and covariantly definable in terms of the elements of M_1 and the geometrical objects associated with M_1 are uniquely and covariantly definable in terms of M_2, and vice versa.*

It is this criterion that GNG and NG (allegedly) fail, despite being empirically equivalent. The reason is that, although it is always possible given a model M_1 of NG to uniquely and

[24] In other words, the ability to translate from each theory to the other does not imply definitional equivalence. See Andréka et al. (2005) for an example.

[25] Actually, Glymour states his condition (which is really only part (2) of condition 1) as a necessary condition for theoretical equivalence. He is mute on whether it is also sufficient, or whether there is some additional condition that he thinks should be satisfied. It seems to me, however, that in any context in which definitional equivalence is a natural necessary condition, it is also a natural sufficient condition—at least when combined with a precise notion of empirical equivalence. In other words, if the sense of equivalence one wants is synonymy of the sort that definitional equivalence captures, then it is hard to see what else, aside from definitional equivalence and empirical equivalence, one could ask for. As I said above, however, there may well be circumstances under which one would want a different condition altogether—indeed, I will presently argue that the equivalence of GNG and NG is best characterized by a different condition.

covariantly define a model M_2 of GNG, it is *not* possible to go in the other direction: given M_2, there are a continuum of models of NG that yield the same model of GNG, and so there is no hope of uniquely picking out one of them in the necessary way. Thus, it would seem, GNG and NG are not synonymous.

4. Another sense of synonymy?

I will presently argue that condition 1 does not capture the sense in which two clearly synonymous formulations of electromagnetism are equivalent. For simplicity, I will limit attention to source-free electromagnetism in a fixed background of Minkowski spacetime.[26] Source-free electromagnetism describes the behavior of electromagnetic fields, which are represented by a smooth, antisymmetric tensor field $F_{ab}(= F_{[ab]})$ on M, called the *Faraday tensor*.[27] This field satisfies Maxwell's equations, which in the present language can be expressed compactly as

$$\nabla_{[a}F_{bc]} = \mathbf{0} \tag{4.1a}$$

$$\nabla_{a}F^{ab} = \mathbf{0}. \tag{4.1b}$$

For present purposes, I will take for granted that the empirical content of this theory amounts to a specification of F_{ab} at every point of spacetime.

In many contexts, the field F_{ab} is cumbersome to work with directly. Instead, physicists work with a vector field, A^a, called the 4-vector potential. The 4-vector potential bears a simple relationship to the Faraday tensor, given by

$$F_{ab} = \nabla_{[a}A_{b]}. \tag{4.2}$$

[26] Minkowski spacetime is a (fixed) relativistic spacetime (M, η_{ab}) characterized by three features: M is $\mathbb{R}^4$, η_{ab} is a flat Lorentzian metric, and the spacetime is geodesically complete.

[27] The notation $[\cdot]$, applied to tensor indices, refers to anti-symmetrization over these indices.

When F_{ab} is defined in terms of a vector potential A^a, Eq. (4.1a) is automatically satisfied. (Conversely, assuming that an arbitrary F_{ab} field satisfies Eq. (4.1a) guarantees that there exists some vector field A^a such that $F_{ab} = \nabla_{[a}A_{b]}$.)[28] Meanwhile, we can write Eq. (4.1b) in terms of A^a as

$$\nabla_a \nabla^a A^b = \nabla^b \nabla_a A^a. \tag{4.3}$$

Eq. (4.3) can be understood as the field equation governing this second formulation of electromagnetism. Once again, the empirical content of the theory should be understood to be its specification of F_{ab} at every point, as determined using Eq. (4.2).

These two ways of describing the theory—in terms of F_{ab} and in terms of A^a—give rise to two ways of characterizing the models of the theory. In one, models can be thought of as a specification of a particular tensor field F_{ab}, satisfying Maxwell's equations. We can write these as ordered triples (M, η_{ab}, F_{ab}), where the first two elements of the triplet refer to the background spacetime. Call these the models of EM$_1$. In the second approach, models can be thought of as a specification of a particular vector field A^a, satisfying Eq. (4.3). These can be written as ordered triplets (M, η_{ab}, A^a). They are the models of EM$_2$. Expressed in this way, EM$_1$ and EM$_2$ can be conceived of as two different theories—though they are certainly empirically equivalent, since given any model of EM$_1$, there exists a model of EM$_2$ that assigns the same F_{ab} tensor to each point of spacetime, and vice versa.

Understood as two empirically equivalent theories of covariant objects on a manifold—the two versions of electromagnetism are amenable to analysis by condition 1. And so one might ask, are EM$_1$ and EM$_2$ equivalent under this criterion? Evidently not.

Proposition 4.1. *EM$_1$ and EM$_2$ are not theoretically equivalent by condition 1.*

Proof. Given any model of EM$_2$, (M, g_{ab}, A^a), one can always (covariantly) define a unique model of EM$_1$ by Eq. (4.2). Conversely, given any model of EM$_1$, (M, g_{ab}, F_{ab}), there exists

[28] Briefly, if $F_{ab} = \nabla_{[a}A_{b]}$, then F_{ab} is an exact 2-form, so it is automatically closed. Conversely, if F_{ab} is closed, and the background manifold is simply connected (as $\mathbb{R}^4$ is), then F_{ab} is exact—so there exists some field A^a such that $F_{ab} = \nabla_{[a}A_{b]}$.

a vector field A^a such that (M, g_{ab}, A^a) is a model of EM$_2$, and moreover, Eq. (4.2) is satisfied. However, this second translation is not unique. Pick a model of EM$_2$, (M, η_{ab}, A^a), and consider a map on this model that takes $A^a \mapsto A'^a = A^a + \nabla^a \psi$, for some smooth scalar field ψ. This new triplet, (M, η_{ab}, A'^a), is itself a model of EM$_2$, since it satisfies Eq. (4.3) (for any choice of ψ). Now consider how the Faraday tensor determined by A^a behaves under the transformation. We have,

$$F_{ab} \mapsto F'_{ab} = \nabla_{[a}(A_{b]} + \nabla_{b]}\psi) = \nabla_{[a}A_{b]} + \nabla_{[a}\nabla_{b]}\psi = F_{ab},$$

where the last equation holds because the derivative operator is torsion-free (and so $\nabla_a \nabla_b \psi = \nabla_b \nabla_a \psi$ for *any* scalar field). Thus, given a model of EM$_1$, there exist a continuum of models of EM$_2$ that agree with regard to the electromagnetic field, and so EM$_1$ and EM$_2$ are not theoretically equivalent by condition 1. □

What should we make of this result? Surely Prop. 4.1 indicates a sense in which EM$_1$ and EM$_2$ are inequivalent—namely they are inequivalent by the standard set by condition 1. But this might give one pause. EM$_1$ and EM$_2$ are supposed to be different formulations of the same theory. There is a strong sense in which they say precisely the same thing about the world, at least on their standard interpretations. The reason concerns the nature of the relationship between the models of EM$_2$. The transformation that takes A^a to A'^a is often called a "gauge transformation" (and electromagnetism, a "gauge theory"). Gauge transformations are considered a special sort of transformation between models of physical theories, because they only affect degrees of freedom in a physical system that are interpreted to be redundant, or unphysical. For this reason, when one works with EM$_2$, it is standard to identify models that differ only by a gauge transformation. This means that if a model of EM$_1$ and a model of EM$_2$ agree regarding their F_{ab} fields, then they agree, period. The vector potential is not supposed to be a real feature of the world; the electromagnetic field derived from it is.[29]

It seems to me that there is a clear and robust sense—indeed, the sense that one would have had in mind to begin with—in which two theories should be understood as synonymous if, on their standard interpretations, they differ only with regard to features that, by the lights of the theories themselves, have no physical content. And EM_1 and EM_2 are examples of theories that do just that. Thus, while perhaps there are situations in which condition 1 captures what one means by theory synonymy, at least for electromagnetism it would appear that it does not. At very least, there is *a* sense of synonymy that is salient and interesting, but which condition 1 apparently misses.

There are two reasons why EM_1 and EM_2 fail to be equivalent by condition 1. The first problem concerns the nature of gauge symmetry, which explicitly identifies models of a theory that differ only with regard to structure that does not have physical meaning—i.e., only with regard to mathematical auxiliary structure. Condition 1 cannot accommodate a situation in which a translation fails to be unique only because there are a number of *equivalent* models to translate a given model into. Really one wants to allow theories to be equivalent if their models are uniquely intertranslatable up to some antecedent notion of model equivalence.

The second difficulty is more pernicious. It concerns "covariantly definability". On the natural interpretation of covariant definability, the Faraday tensor F_{ab} is always definable in terms of a vector potential, via Eq. (4.2). But in general, there is no way to define a vector potential in terms of a covariant formula involving the Faraday tensor. Rather, one has general existence results that guarantee that for any Faraday tensor satisfying Eq. (4.1a), an associated vector potential must exist.[30] This problem is more difficult to resolve than

[29] Once one moves to quantum mechanics, the status of the vector potential changes (c.f. the Aharanov-Bohm effect). But for present purposes, in classical electromagnetism, the interpretation of the vector potential given here is standard. However, see Belot (1998) and Healey (2007).

[30] These two problems really are distinct. On the one hand, one can imagine a situation where multiple

the first because if one were to drop the requirement of explicit covariant definability from condition 1 altogether, the condition would appear to collapse into empirical equivalence. Perhaps in some situations, one should follow the positivists and take empirical equivalence as a standard of theoretical equivalence. But in *this* case, that seems unsatisfactory: EM_1 and EM_2 are not merely empirically equivalent theories, they are the same theory, expressed in different terms. One would like to be able to articulate what this stronger sense of equivalence amounts to.

5. An alternative criterion of equivalence

Thus far, I have introduced a criterion of theoretical equivalence and argued that it fails to capture the sense in which EM_1 and EM_2 are synonymous. In the present section, I will present a criterion of equivalence that does capture the sense in which EM_1 and EM_2 are synonymous. The condition I have in mind is most naturally stated within the setting of category theory.[31] I hope the motivation for bringing in this new mathematical machinery will become clear as the section progresses, but perhaps it will be useful to make a few remarks up front. As should be clear from the conclusion of the last section, in order to adequately represent a gauge theory, one needs information both about the models of the theory and also about how those models relate to one another. In the present case, that information amounts to a specification of which models of a theory should be taken to be physically equivalent to one another. Since the equivalences necessary to adequately represent a gauge theory can be thought of as a privileged collection of maps between models, category theory is a natural mathematical setting for the present discussion. More

fields equivalent for the purposes at hand may be explicitly defined, depending on the choice of some fixed (non-unique) background field. On the other hand, one can imagine a situation where one has an existence result that, under certain circumstances, some field must exist, and yet be unable to define that field explicitly in terms of whatever background fields one has access to.

[31] I assume basic knowledge of category theory. For background, see Awodey (2006), Mac Lane (1998), or Borceux (2008).

importantly, category theory provides a way of representing what, in addition to empirical equivalence, models of EM_1 and EM_2 have in common. The alternative characterization of theoretical equivalence I provide will trade on a notion of implicit definability, as opposed to the explicit definability required by condition 1.[32]

I will first reconstruct condition 1 in category theoretic terms. As a start, note that condition 1 is stated as a relationship between the models of each theory. This means that to apply condition 1, one begins by representing a theory by a collection of models of that theory.[33] To use the example of electromagnetism from the previous section, EM_1 and EM_2 are represented by the collections of triples of the form (M, η_{ab}, F_{ab}) and the triples of the form (M, η_{ab}, A^a), respectively. (As a matter of notation, I will use italic type—EM_1 and EM_2—to refer to the collections of models, and for all practical purposes I will treat them as sets.) Condition 1, then, is the requirement that there exists a pair of maps $F : EM_1 \to EM_2$ and $G : EM_2 :\to EM_1$, satisfying certain conditions: namely that (1) F and G map models of one theory to empirically equivalent models of the other theory, (2) F and G map models to models in such a way that the elements of the destination models can be covariantly defined in terms of the elements of the source models, and (3) $G \circ F$ and $F \circ G$ act as the identity on EM_1 and EM_2, respectively.

To translate this into category theoretic terms, we first represent the two varieties of electromagnetism as categories, $\mathcal{EM}_1$ and $\mathcal{EM}_2$.[34] The objects of $\mathcal{EM}_1$ and $\mathcal{EM}_2$ are just

[32] Once again, it is likely possible to recover this discussion in terms of models and equivalence classes of models. Nonetheless, the category theoretic setting makes the difference between the two conditions much clearer, and avoids the charge that the modified condition I propose is *ad hoc*.

[33] This way of thinking about theories is often called the "semantic view". I do not mean to lump Glymour in with the strictest adherents to this view; all I meant to say is that since his criterion is expressed in terms of relations between individual models of a theory, he is in effect using a theory's models to represent the theory as a whole. The remarks in the present section can be construed as an argument against the view that a theory can be represented (simply) by a collection of models, without any additional structure. For more on this point, see Halvorson (2012) and Halvorson and Weatherall (2012).

[34] In what follows, I will sometimes refer to specific categories as theories. But I should be clear, I mean categories that are representations of theories. In particular, while a theory may be representable as a category of its models, a category need not represent a theory.

the elements of EM_1 and EM_2, respectively. But now there is an immediate question that did not come up when describing EM_1 and EM_2, regarding the arrows in the categories $\mathcal{EM}_1$ and $\mathcal{EM}_2$. At first pass, one might include only the identity arrows.[35] But a moment's reflection reveals that there are a wide variety of privileged maps between models of electromagnetism (in either formulation)—namely, maps that preserve the "physical structure" of a model, in the sense that two models related by such a map are physically equivalent.[36]

For instance, consider a model $(M, \eta_{ab}, F_{ab}) \in EM_1$ and an isometry φ from Minkwoski spacetime to itself.[37] Then the triplet $(M, \eta_{ab}, \varphi^*(F_{ab}))$ is also an object of $\mathcal{EM}_1$ (or an element of EM_1), since the spacetime is Minkwoski spacetime, and $\varphi^*(F_{ab})$ is an anti-symmetric tensor field satisfying Maxwell's equations. Unless φ is the identity map, (M, η_{ab}, F_{ab}) will not be equal $(M, \eta_{ab}, \varphi^*(F_{ab}))$. However, these two models are naturally understood to represent precisely the same physical situation—indeed, in general one can understand the diffeomorphism φ as implementing a change of coordinates, and the push-forward map on F_{ab} simply determines the field in the new coordinate system. When one ordinarily describes models of electromagnetism, the fact that some models represent the same physical situation in this way is implicit. But in the present context, we have a way of making these relations between models explicit, or perhaps better, of representing the class of structure-preserving transformations as part of the physical theory.

To be precise, I will define $\mathcal{EM}_1$ and $\mathcal{EM}_2$ such that there will be a map f in the collection of arrows between models (M, η_{ab}, F_{ab}) and (M, η_{ab}, F'_{ab}) of EM_1 (or EM_2, *mutatis mutandis*) if there exists an isometry $\psi : M \to M$ such that $\psi^*(F_{ab}) = F'_{ab}$. When this

[35] Indeed, to recover condition 1 exactly, perhaps it would appropriate to work with these impoverished categories.

[36] One might be tempted to call such maps "symmetries" of the theories, but as Gordon Belot (2012) has pointed out, there are several ways in which the term "symmetry" is used so I will avoid it. The maps I have in mind the ones that take models to physically equivalent models, and not necessarily ones that correspond to other notions of symmetry.

[37] Let (M, η_{ab}) be Minkwoski spacetime. A Minkowski spacetime isometry is a diffeomorphism $\varphi : M \to M$ such that $\varphi^*(\eta_{ab}) = \eta_{ab}$, where φ is the push-forward map associated with φ. These maps correspond to the Poincaré group, consisting of timelike and spacelike translations, Lorentz boosts, and spatial rotations.

occurs, I will say that f is "generated" by φ. Note, too, that this relation is symmetric, and indeed, $\psi^{-1} : M' \to M$ will always generate an arrow in the opposite direction. (In other words, the maps I have just described are invertible—in category theoretic terms, these morphisms are *isomorphisms.*)

A virtue of using the arrows of the categories to represent maps between physically equivalent models is that it provides a natural way of representing gauge transformations in EM_2: they are simply another class of arrows between equivalent models. Thus in addition to the maps that preserve M, η_{ab}, and A^a in the sense described above, the arrows of $\mathcal{EM}_2$ should also include the collection of arrows corresponding to gauge transformations between models of EM_2 (as well as all compositions of gauge transformations and isometry-generated maps). Note that these models are such that if a map exists between two of them, it must be the unique map between those models, since a map between these triplets can only act in one way. Including gauge transformations with to the morphisms of $\mathcal{EM}_2'$ demonstrates how specifying a collection of maps can provide information about what the physically relevant structure of a model is. This information is not contained in a specification of a single model of EM_2, or even by specifying all of the models. But it is encoded once one indicates which models are physically equivalent.

The next step to translating Glymour's condition concerns the maps F and G described above. These now have a natural interpretation as a special kind of functor between $\mathcal{EM}_1$ and $\mathcal{EM}_2$, which I will call a "translation functor".

Definition 5.1. *Let $\mathcal{T}_1$ and $\mathcal{T}_2$ be categories whose objects are models of a physical theory, expressed in terms of tensor fields on a manifold. A* translation functor *between $\mathcal{T}_1$ and $\mathcal{T}_2$ is a functor $F : \mathcal{T}_1 \to \mathcal{T}_2$ that maps objects in $\mathcal{T}_1$ to objects in $\mathcal{T}_2$ whose elements can be covariantly defined in terms of the elements of the source object.*

In the present language, then, part of Glymour's criterion for equivalence will be the existence of a suitable pair of translation functors.

Condition 1 also requires that the two theories be empirically equivalent, in the sense that for every model of one theory there is a model of the other that makes the same predictions, and vice versa. This condition can also be absorbed into the formalism by considering an additional category $\mathcal{P}(T)$ whose objects are the possible "predictions" of a theory T. Given such a category, the relationship between a theory and its predictions can be represented as a functor $P : \mathcal{T} \rightarrow \mathcal{P}(T)$ that takes models of a theory to that model's predictions.[38] The very existence of a category of predictions assumes a great deal about a theory. Moreover, even in cases like the present one, where the predictions of the theories are reasonably clear, there is quite a bit of flexibility in how one represents the category of predictions. To be concrete, in what follows I will understand $\mathcal{P}(EM)$ to be the *skeleton* of $\mathcal{EM}_1$, which is the category whose objects are representative elements of the equivalence classes generated by the arrows of $\mathcal{EM}_1$, and whose arrows are just the identity map.[39] The functors $P_{EM_1} : \mathcal{EM}_1 \rightarrow \mathcal{P}(EM)$ and $P_{EM_2} : \mathcal{EM}_2 \rightarrow \mathcal{P}(EM)$ then take models to the physical situation they represent in the obvious way, and take all arrows to identity arrows.

We now have the apparatus to state condition 1 as a relation between categories, using the notion of an isomorphism between categories.

Definition 5.2. *Two categories $\mathcal{C}$ and $\mathcal{D}$ are* isomorphic *if and only if there exist functors $F : \mathcal{C} \rightarrow \mathcal{D}$ and $G : \mathcal{D} \rightarrow \mathcal{C}$ such that $F \circ G = 1_{\mathcal{D}}$ and $G \circ F = 1_{\mathcal{C}}$, where $1_{\mathcal{C}}$ and $1_{\mathcal{D}}$ are the identity functors on $\mathcal{C}$ and $\mathcal{D}$, respectively.*

Condition 1′. *Theories $\mathcal{T}_1$ and $\mathcal{T}_2$ with a common category of predictions $\mathcal{P}(T)$ are theoretically equivalent if and only if (1) there exists a pair of translation functors $F : \mathcal{T}_1 \rightarrow \mathcal{T}_2$ and $G : \mathcal{T}_2 \rightarrow \mathcal{T}_1$ that together realize a categorical isomorphism between $\mathcal{T}_1$ and $\mathcal{T}_2$, and (2) the following diagram commutes (in both directions):*

[38] In some cases, one might represent a theory as a triple consisting of a category of models, a category of predictions, and a functor between them.

[39] A skeleton category bears a close relationship to a quotient set; one can think of the objects of the prediction category thus defined as the various physical situations represented by the equivalence classes of models in each category.

$$\begin{tikzcd}
\mathcal{T}_1 \arrow[rr, bend right, "F"'] \arrow[dr, "P_1"'] & & \mathcal{T}_2 \arrow[ll, bend right, "G"'] \arrow[dl, "P_2"] \\
& \mathcal{P}(T) &
\end{tikzcd}$$

This restatement really does capture the content of condition 1: in effect, it requires that every model of $\mathcal{T}_1$ can be explicitly translated into a model of $\mathcal{T}_2$ that makes the same predictions, and moreover the process can be reversed.

Proposition 5.3. *$\mathcal{EM}_1$ and $\mathcal{EM}_2$ are not equivalent by condition 1′.*

I will suppress the proof, as it is essentially the same as that of Prop. 4.1.

Thus far, I have simply recovered condition 1 and Prop. 4.1 in a new setting. But now that I have translated these results, I have the resources available to formulate an alternative condition that avoids the problems described at the end of the last section. In the present language, the first problem is that condition 1′ requires that the two categories be isomorphic. What we really want is a notion of isomorphism of categories "up to model equivalence". This idea can be made precise using an *equivalence of categories.*[40]

Definition 5.4. *Two categories $\mathcal{C}$ and $\mathcal{D}$ are* equivalent *if and only if there exist functors $F : \mathcal{C} \to \mathcal{D}$ and $G : \mathcal{D} \to \mathcal{C}$ such that (1) for every object $A \in \mathcal{C}$, there is an isomorphism $\eta_A \in \hom(G \circ F(A), A)$ such that for every object $B \in \mathcal{C}$ and every morphism $f \in \hom(A, B)$, the following diagram commutes:*

[40] This definition may call for some explanation. Two categories are equivalent if there are a pair of functors that are "almost inverses", in the sense that if you apply the first and then apply the second, you come back to an object that is (1) isomorphic (by the standard of isomorphism determined by the category in question) and (2) whose arrows behave in the same way as the original object's arrows. To take a concrete example, consider the simple categories $\mathcal{C}$, which consists of two objects A and B with their identities and a pair of isomorphisms $f \in \hom(A, B)$ and $g \in \hom(B, A)$, and $\mathcal{D}$, which consists of a single object C and its identity. These two categories can be represented by the following pictures.

$$\begin{tikzcd}
\bullet_A \arrow[loop left, "1_A"] \arrow[r, bend left, "f"] & B\bullet \arrow[l, bend left, "g"] \arrow[loop right, "1_B"] & & & \bullet_C \arrow[loop left, "1_C"]
\end{tikzcd}$$

These two categories are equivalent, with the equivalence realized by the following two functors: $F : \mathcal{C} \to \mathcal{D}$, which maps A and B to C, and maps all of the arrows to 1_C, and $G : \mathcal{D} \to C$, which maps C to A and 1_C to 1_A.

$$\begin{tikzcd}
G \circ F(A) \arrow[r, "G\circ F(f)"] \arrow[d, "\eta_A"'] & G \circ F(B) \arrow[d, "\eta_B"] \\
A \arrow[r, "f"'] & B
\end{tikzcd},$$

and conversely, (2) for every object $C \in \mathcal{D}$, there is an isomorphism $\epsilon_C \in \hom(F \circ G(C), C)$ such that for every object $D \in \mathcal{D}$ and every morphism $g \in \hom(C, D)$, the corresponding diagram,

$$\begin{tikzcd}
F \circ G(C) \arrow[r, "F\circ G(g)"] \arrow[d, "\epsilon_C"'] & F \circ G(D) \arrow[d, "\epsilon_D"] \\
C \arrow[r, "g"'] & D
\end{tikzcd},$$

commutes.

For essentially all mathematical intents and purposes, two categories are considered "the same" if they are equivalent, even if they not necessarily isomorphic. Equivalence of categories is precisely the required notion of equivalence "up to model equivalence".

The second problem identified at the end of the last section is the requirement that there exist *translation* functors that realize the categorical isomorphism. At the end of the previous section, I argued that one could not simply drop this requirement, since without it condition 1 would reduce to empirical equivalence. But in the present setting, that is not the case. The reason is that the categories $\mathcal{EM}_1$ and $\mathcal{EM}_2$ contain far more information than the bare sets EM_1 and EM_2, encoded in the arrows of the categories. A categorical equivalence between the categories $\mathcal{EM}'_1$ and $\mathcal{EM}'_2$ preserves this additional information regarding the arrows of the two theories.

Why should this matter? This relationship between the theories can be understood in terms of implicit definability. One way of thinking about implicit definability quite generally is that an object (or a relation, or a function) is implicitly definable in terms of other objects just in case the first object is invariant under precisely the same class of transformations as the objects in terms of which it is to be defined. This notion can be made perfectly precise in first-order logic, at least for complete theories, by way of Svenonius' theorem.[41] In first-order logic, of course, it is well known that implicit definability and explicit definability collapse into one another—this is the content of Beth's celebrated definability theorem.[42] In other logics, however, including full second-order logic, Beth's theorem is known to fail, which means that implicit and explicit definability pull apart.[43] In the present context, we are not working in a precise logical setting, but nonetheless we can draw some insight by analogy: even in contexts where invariance under a class of transformations is not equivalent to explicit definability, such invariance may nonetheless be interpreted as providing *some* notion of (non-explicit) definability. In the present case, one can understand the thing being implicitly defined to be the physical configuration represented by an equivalence class of models, since this is what is preserved under the relevant class of transformations.[44]

These considerations suggest the following alternative to conditions 1 and 1′.

Condition 2. *Theories $\mathcal{T}_1$ and $\mathcal{T}_2$ with a common category of predictions $\mathcal{P}(T)$ are theoretically equivalent if and only if (1) there exists a pair of functors $F : \mathcal{T}_1 \rightarrow \mathcal{T}_2$ and $G : \mathcal{T}_2 \rightarrow \mathcal{T}_1$ that together realize a categorical equivalence between $\mathcal{T}_1$ and $\mathcal{T}_2$, and (2) the following diagram commutes (in both directions):*

$$\begin{tikzcd}
\mathcal{T}_1 \arrow[rr, bend right, "F"'] \arrow[dr, "P_1"'] & & \mathcal{T}_2 . \arrow[ll, bend right, "G"'] \arrow[dl, "P_2"] \\
& \mathcal{P}(T) &
\end{tikzcd}$$

[41] See Hodges (1993, Corollary 10.5.2) and surrounding discussion.

[42] See Hodges (1993, Theorem 6.6.4) and surrounding discussion.

[43] See Paulos (1976), Burgess (1977), and Makowsky and Shelah (1979). A notion of definability within physical theories that trades on a similar analogy with Svenonius' theorem is described by Halvorson and Swanson (2012).

[44] In particular, I do not mean to say that a vector potential can be implicitly defined in the way described above. What is defined is the invariant structure under the equivalence relations in the categories. This, I claim, is what is shared between two categorically equivalent theories.

By this criterion, EM_1 and EM_2 *are* equivalent.

Proposition 5.5. *$\mathcal{EM}_1'$ and $\mathcal{EM}_2'$ are equivalent by condition 2.*

Proof. To prove that any two categories $\mathcal{C}_1$ and $\mathcal{C}_2$ are equivalent, it suffices to show that there exists a functor $F : \mathcal{C}_1 \to \mathcal{C}_2$ that is (1) full and faithful (i.e., for any $A, B \in \mathcal{C}_1$, the map $F : \hom(A, B) \to \hom(F(A), F(B))$ is bijective) and (2) essentially surjective (i.e., for any $A \in \mathcal{C}_2$, there is an $A' \in \mathcal{C}_1$ such that there exists an isomorphism $f \in \hom(A, F(A'))$). We can define such a functor $F : \mathcal{EM}_2 \to \mathcal{EM}_1$ as follows. For any model $A = (M, \eta_{ab}, A^a) \in \mathcal{EM}_2$, set $F(A) = (M, \eta_{ab}, F_{ab}) \in \mathcal{EM}_1$, where $F_{ab} = \nabla_{[a}A_{b]}$. Now consider any pair of models $A = (M, \eta_{ab}, A^a)$ and $A' = (M, \eta_{ab}, A'^a)$ in $\mathcal{EM}_2$ for which $\hom(A, A')$ is non-empty. Take $f \in \hom(A, A')$. If $F(A) = F(A')$, set $F(f) = 1_{F(A_1)}$. Otherwise, there must be an isometry φ such that $\varphi^*(A^a) = A'^a$. I claim that φ generates a morphism $g \in \hom(F(A), F(A'))$. Since φ is already a Minkowski spacetime isometry, we only need to show that $\varphi^*(F_{ab}) = F'_{ab}$ (where $F'_{ab} = \nabla'_{[a}A_{b]}$). It does: $\varphi^*(F_{ab}) = \varphi^*(\nabla_{[a}A_{b]}) = \nabla'_{[a}\varphi^*(A_{b]}) = \nabla'_{[a}A'_{b]} = F'_{ab}$. So set $F(f) = g$. Note that F thus defined is a functor, since it inherits the composition rule from the properties of diffeomorphisms. Moreover, F is surjective on objects (because every Faraday tensor has a vector potential that maps to it), which means it is automatically essentially surjective.

It remains to show that F is full and faithful. First, for any $A, A' \in \mathcal{EM}_2$, the restriction of F to $\hom(A, A')$ is clearly injective, since either $\hom(A, A') = \emptyset$ or else it has a unique element. Now suppose that there is some $g \in \hom(F(A), F(A'))$ such that, for all $f \in \hom(A, A')$, $F(f) \neq g$. If $\hom(F(A), F(A')) = \emptyset$, we are finished, so assume that $\hom(F(A), F(A'))$ is non-empty. If $F(A) = F(A')$, then g must be the identity morphism, and either $A = A'$ or A and A' must be related by a gauge transformation. In either case, there is a map $f \in \hom(A, A')$ such that $F(f) = 1_{F(A)}$. So it must be that $F(A) \neq F(A')$. It follows that there must be some diffeomorphism $\varphi : M \to M'$ that generates g. But I claim that φ also generates a map between A and A'. To see this, note that φ is already a Minkowski spacetime isometry, and moreover, since $\varphi^*(F_{ab}) = F'_{ab}$, $\varphi^*(\nabla_{[a}A_{b]}) = \nabla'_{[a}A'_{b]}$. This last equation implies that either (1) $\varphi^*(A_a) = A'_a$ or (2) $\varphi^*(A_a)$ and A'_a are related by a gauge transformation (because $\varphi^*(A^a)$ and A'^a have the same exterior derivative). In the first case, there is a map $f \in \hom(A, A')$ such that $F(f) = g$, by the definition of F—namely the map f generated by φ. And in the second case, there is a model A'' such that A_a and A''_a are related by a gauge transformation $h \in \hom(A, A'')$, and such that $\varphi^*(A''_a) = A'_a$. Thus there is a map $h' \in \hom(A'', A')$ that is generated by φ. Moreover, $F(A'') = F(A)$ and $F(h') \in \hom(F(A), F(A'))$ must be g, since it is generated by φ. And so $F(h' \circ h) = F(h') \circ F(h) = g \circ 1_A = g$. Thus F restricted to $\hom(A, A')$ is surjective for any $A, A' \in \mathcal{EM}_2$, and F is full and faithful.

We have now shown that F is one half of a categorical equivalence between $\mathcal{EM}_1$ and $\mathcal{EM}_2$. Moreover, $P_{EM_1} \circ F = P_{EM_2}$, since F takes objects of $\mathcal{EM}_2$ to objects of $\mathcal{EM}_1$ that yield precisely the same predictions. It remains to show only that there exists a functor G that forms the other half of the equivalence and such that $P_{EM_2} \circ G = P_{EM_1}$. But this is easily done: for each $B = (M, \eta_{ab}, F_{ab}) \in \mathcal{EM}_1$, simply take $G(B) = (M, \eta_{ab}, A^a)$, where A^a is some vector potential for which $F_{ab} = \nabla_{[a}A_{b]}$, and let the action of G on arrows simply be the inverse of the action of F on arrows. Thus F and G realize a categorical equivalence, and moreover, $P_{EM_2} \circ G = P_{EM_1}$. □

6. Are NG and GNG theoretically equivalent?

We can now revisit the question at the heart of the present paper. Are NG and GNG theoretically equivalent by condition 2? One first needs to say what the categories associated with NG and GNG are going to be. In the case of the category $\mathcal{NG}$, the objects will be ordered sextuples $(M, t_a, h^{ab}, \nabla, \varphi, \rho)$, with ∇ flat, that satisfy Poisson's equation (and the curvature conditions (CC2) and (CC3)). In the category $\mathcal{GNG}$, meanwhile, the objects will be ordered quintuples $(M, t_a, h^{ab}, \nabla, \rho)$ that satisfy the geometrized version of

Poisson's equation. But what about the arrows between the models? For $\mathcal{GNG}$, the answer seems reasonably clear. As with electromagnetism, one should take the arrows to represent transformations between physically equivalent models. Two models $(M, t_a, h^{ab}, \nabla, \rho)$ and $(M', t'_a, h'^{ab}, \nabla', \rho')$ are physically equivalent just in case there exists a diffeomorphism $\psi : M \to M'$ such that $\psi^*(t_a) = t'_a$, $\psi^*(h^{ab}) = h'^{ab}$, $\rho = \rho' \circ \psi$, and for any geodesic γ of ∇, $\gamma \circ \psi$ is a geodesic of ∇'.[45]

The situation for $\mathcal{NG}$ is not quite so clear, however. One certainly wants to include arrows of the sort I have suggested we include in $\mathcal{GNG}$, corresponding to the explicit structure preserving maps. But are these the only arrows in the category? It depends. There are two choices, corresponding to two "natural" ways of construing NG:

Option 1. One takes models of NG that differ with regard to the gravitational field to be distinct.

Option 2. One takes models of NG whose gravitational field and derivative operators are related by the transformation $\varphi \mapsto \varphi + \psi$ and $\nabla \mapsto \nabla' = (\nabla, t_b t_c \nabla^a \psi)$, for any smooth ψ satisfying $\nabla^a \nabla^b \psi = \mathbf{0}$, to be equivalent.[46]

In other words, in the second case one takes the gravitational potential to be a gauge quantity, much like the vector potential in electromagnetism. In the first case, one does not. These two options correspond to two different categories. In one case, call it $\mathcal{NG}_1$ (and the associated theory, NG_1), one does not add any arrows to the category $\mathcal{NG}$ as already described. The only structure-preserving maps are the ones generated by diffeomorphisms. This category corresponds to option 1. To represent option 2, on the other hand, one adds an additional collection of arrows between objects related by the transformation indicated.

[45] This condition makes sense because a derivative operator is fully characterized by its geodesics (Malament, 2012)[Prop. 1.7.8].

[46] Note that, since *all* of the derivative operators considered in NG and GNG agree once one raises their index, one can characterize the gauge transformation with regard to any of them without ambiguity.

Call this category $\mathcal{NG}_2$ (and its associated theory NG_2). These additional arrows play the same role in $\mathcal{NG}_2$ that the arrows between gauge-equivalent vector potentials play in $\mathcal{EM}_2$.

What considerations might lead one to adopt one view of the theory over the other? I believe, though I am not sure, that the first option aligns better with how physicists have thought of Newtonian gravitation historically. And moreover, in the presence of certain boundary conditions—for instance, the assumption that the gravitational field vanishes at spatial infinity, as is natural when considering matter distributions with spatially compact support, such as the solar system—there is always a unique choice of gravitational field/derivative operator for a given matter distribution. If one has reason to think that a particular choice of potential/derivative operator is privileged, there is little to recommend identifying the privileged choice with other, apparently less physical fields.

That said, there are systems in which option 1 leads to problems. For instance, in certain cosmological situations, such as in spacetimes with homogeneous and isotropic matter distributions, one would expect the gravitational field to be non-vanishing everywhere, including at spatial infinity. And indeed, under such circumstances, option 1 generates explicit contradictions, where one can derive that the gravitational potential at any point takes on multiple values. But if one takes these different choices of gravitational field to be equivalent, then the apparent contradiction dissolves.[47] From this point of view, while there may be systems in which a particular choice of potential/derivative operator is more convenient than others, one should nonetheless adopt option 2 in general. Such arguments strike me as compelling, and I tend to agree with the conclusion that option 2 is preferable.

But I will not argue further for this thesis, and for the purposes of the present paper, I will remain agnostic about which way of understanding NG is preferable. Rather, the point is simply to remark that once one has distinguished these two possibilities, it is possible to ask the question with which I began this paper with additional care. Really, we have

[47] For more on this, see the debate between John Norton (1992, 1995) and David Malament (1995).

a variety of questions at hand: we have three categories ($\mathcal{NG}_1$, $\mathcal{NG}_2$, and $\mathcal{GNG}$) and two conditions (conditions $1'$ and 2). Are any of these theories pairwise equivalent by either criterion?[48] None of these theories are equivalent by condition $1'$. Moreover, $\mathcal{NG}_1$ is not equivalent to either $\mathcal{GNG}$ or $\mathcal{NG}_2$ by condition 2. But $\mathcal{GNG}$ and $\mathcal{NG}_2$ *are* theoretically equivalent by condition 2.[49] (The situation is summarized by table 1.)

Proposition 6.1. *$\mathcal{NG}_2$ and $\mathcal{GNG}$ are theoretically equivalent by condition 2.*

The argument is essentially the same as the proof of Prop. 5.5. I will exhibit the full, faithful, and essentially surjective functor $F : \mathcal{NG}_2 \to \mathcal{GNG}$ and suppress the remaining details. This functor can be defined as the one that takes objects $(M, t_a, h^{ab}, \nabla, \varphi, \rho) \in \mathcal{NG}_2$ to objects $(M, t_a, h^{ab}, \overset{g}{\nabla}, \rho)$ where $\overset{g}{\nabla} = (\nabla, -t_b t_c \nabla^a \varphi)$. By Theorem 2.2, we know that this destination object must be in $\mathcal{GNG}$. Now consider any $A, A' \in \mathcal{NG}$ such that $\hom(A, A')$ is non-empty, and any $f \in \hom(A, A')$. If $F(A) = F(A')$, then we can set $F(f) = 1_{F(A)}$. Otherwise, f is generated by some diffeomorphism $\psi : M \to M'$, which is such that $\psi^*(t_a) = t'_a$, $\psi^*(h^{ab}) = h'^{ab}$, $\varphi = \varphi' \circ \psi$, $\rho = \rho' \circ \psi$, and if $\gamma : I \to M$ is a geodesic of ∇, then $\psi \circ \gamma : I \to M'$ is a geodesic of ∇'. I claim that ψ also generates a map g between $F(A)$ and $F(A')$. We already know that ψ preserves all the structure shared by models of NG and models of GNG; it remains to show that if a curve $\gamma : I \to M$ is a geodesic of $\overset{g}{\nabla}$, then $\psi \circ \gamma$ is a geodesic of $\overset{g}{\nabla}'$. But it must be: if γ is a geodesic of $\overset{g}{\nabla}$, then $\xi^a \overset{g}{\nabla}_a \xi^b = 0$, where ξ^a is the tangent field of γ. We then have $\xi^a(\nabla_a \xi^b + t_a t_n (\nabla^b \varphi) \xi^n) = 0$. Thus $0 = \psi^*(\xi^a(\nabla_a \xi^b + t_a t_n(\nabla^b \varphi)\xi^n)) = \psi^*(\xi^a)(\nabla'_a \psi^*(\xi^a) + t'_a t'_n (\nabla'^b \varphi') \psi^*(\xi^a)) = \psi^*(\xi^a)(\overset{g}{\nabla}'_a \psi^*(\xi^a))$.

[48] Actually, before asking whether any of these are equivalent by conditions $1'$ and 2, I need to define their prediction categories. By analogy to $\mathcal{P}(EM)$, we will take $\mathcal{P}(NG)$ to be the category whose objects are diffeomorphism equivalence classes generated by quintuples $(M, t_a, h^{ab}, \rho, \{\gamma\})$, where $\{\gamma\}$ is a collection of curves representing the trajectories of massive bodies.

[49] There is an observation here that deserves mention. Suppose one restricts attention to the collections of models of NG and GNG in which (1) the matter distribution is restricted to a spatially compact region and (2) the gravitational field (for models of $\mathcal{NG}_1$ and $\mathcal{NG}_2$) vanishes at spatial infinity. Then $\mathcal{NG}_1$, $\mathcal{NG}_2$, and $\mathcal{GNG}$ are *all* theoretically equivalent by condition 2. The reason is that $\mathcal{NG}_1$ and $\mathcal{NG}_2$ collapse into one another, essentially because the conditions I have just described pick out a distinguished gauge. Indeed, the categories are all *isomorphic*. But $\mathcal{GNG}$ is still not equivalent to the others by condition $1'$, because the functor realizing the isomorphism is not a translation functor.

	Condition $1'$	Condition 2
$\mathcal{NG}_1$ and $\mathcal{NG}_2$	Inequivalent	Inequivalent
$\mathcal{NG}_1$ and $\mathcal{GNG}$	Inequivalent	Inequivalent
$\mathcal{NG}_2$ and $\mathcal{GNG}$	Inequivalent	**Equivalent**

Table 1: A summary of the equivalences and inequivalences of NG and GNG, by the standards set by conditions $1'$ and 2.

But since $\psi^*(\xi^a)$ is the tangent field to $\psi \circ \gamma$, it follows that $\psi \circ \gamma$ must be a geodesic relative to $\overset{g}{\nabla}'$. Thus we can set $F(f) = g$. Since diffeomorphisms compose, F is a functor. Moreover, F can be shown to be full, faithful, and essentially surjective by the same set of arguments use in the proof of Prop. 5.5; similarly, it is clear that it commutes with the prediction functors, and that there exists a functor $G : \mathcal{GNG} \to \mathcal{NG}_2$ that with F realizes the equivalence and moreover also commutes with the prediction functors. $\square$

It immediately follows that there exists a (natural) way of construing NG, and a (reasonable) standard of theoretical equivalence such that NG and GNG *are* theoretically equivalent. Moreover, by both this standard of theoretical equivalence and Glymour's standard of theoretical equivalence, the way of construing NG that is theoretically equivalent to GNG is *not* equivalent to another (perhaps also natural) way of construing NG. It is not hard to see what the physical interpretation of this latter inequivalence should be: $\mathcal{NG}_1$, understood as a theory, *reifies* the gravitational field in the sense that it distinguishes spacetimes that differ only with regard to the gravitational field (modulo diffeomorphism); $\mathcal{NG}_2$ does not. In $\mathcal{NG}_2$, the gravitational field is an instrumental quantity, akin to the vector potential, and the choice of a gravitational field/derivative operator pair is a convention. In some cases, there may be a distinguished choice that arises from boundary conditions, but such conditions can at best reflect mathematical convenience, since alternative choices that *fail* to satisfy stated boundary conditions should not be interpreted as attributing particular features to the world. Likewise, the interpretation of the *equivalence* of $\mathcal{GNG}$ and $\mathcal{NG}_2$ is immediate: $\mathcal{NG}_2$ is equivalent to $\mathcal{GNG}$ in just the same sense that $\mathcal{EM}_2$ is equivalent to $\mathcal{EM}_1$: $\mathcal{GNG}$ is a gauge independent formulation of $\mathcal{NG}_2$.

7. The epistemology of geometry and the curvature of spacetime

I have by now made the principal arguments of the paper. These are, in short, that condition 1, does not capture the sense in which EM_1 and EM_2 are synonymous. However, there is a natural alternative to condition 1 that does capture the sense in which EM_1 and EM_2 are synonymous. And by this criterion, GNG and NG are synonymous too, if one takes NG to be a gauge theory in the sense described above. Moreover, condition 2 captures an important physical distinction between the two ways of understanding NG.

There are a few places where one might object to this argument. One might say that it is simply inappropriate to represent a theory with a category in the way I have proposed, either because there is never a satisfactory way to formalize theories or because this particular formalization is lacking. Another kind of objection might be that, even if one accepts the representations of theories proposed here (for some purposes, anyway), condition 2 does not adequately capture any interesting sense of synonymy between theories. I do not agree with either of these objections, but I will not consider them further. For the remainder of this paper, I will suppose that the representation of theories offered here is unobjectionable, and moreover, that condition 2 provides at least a reasonable notion of synonymy that captures some robust sense in which these theories are equivalent. It seems to me that if this is right, there are several philosophical morals to draw.

The first is that this conclusion is not in conflict with at least one of Glymour's principal philosophical claims, regarding the existence of empirically equivalent, theoretically inequivalent theories. This is because even if NG_2 and GNG are theoretically equivalent, NG_1 and GNG are still inequivalent, even by condition 2. Moreover, Glymour's further claim, that GNG is better supported by the empirical evidence even though it is empirically equivalent to NG, is only slightly affected, in that one needs to specify that GNG is only better supported than NG_1 despite their empirical equivalence. But this makes sense: the reason, on Glymour's account, that GNG is better supported than NG is supposed to be that NG makes additional, unsupported ontological claims regarding the existence of a gravitational field. As I have argued, though, a natural way of understanding the difference between NG_1 and NG_2 would be that NG_1 makes the ontological claim that Glymour argues is unsupported, while NG_2 denies it.[50]

There is another purpose to which Glymour puts these arguments, however, that I think *is* affected by the present conclusions. It concerns the epistemology of geometry and the metaphysics of space and time. There is a view, originally due to Poincaré and Reichenbach though also held by others, that one can never know the geometrical properties of spacetime since there always exist empirically equivalent theories that nonetheless differ with regard to (for instance) whether spacetime is curved or flat.[51] These authors conclude that the curvature (or lack thereof) of spacetime is a conventional matter. Glymour argues against conventionalism by pointing out that (on his view of confirmation) the fact that two theories are empirically equivalent does not imply that they are equally well confirmed, and moreover, that one might have better evidence for one member of a collection of empirically equivalent theories over the others. But the present discussion suggests that, at least in some cases, there is another possibility that is not generally considered: theories that attribute apparently *distinct* geometrical properties to the world may nonetheless be synonymous.

Let me make this idea more precise. I have argued that the right way of understanding NG_2 is as a theory that takes the gravitational potential to be a gauge quantity, and the trajectories of bodies given a background matter distribution to be the invariant content of

[50] I should be explicit that I do not mean to endorse any particular views on confirmation here. All I want to say is that if one *were* inclined to hold that GNG is better supported than NG on the grounds Glymour argues for, then the considerations offered in this paper would only nuance that view slightly, in the way specified. Of course, one might have other reasons for rejecting the initial claim altogether.

[51] For a clear and detailed description of the positions that have been defended on the epistemology of geometry in the past, see Sklar (1977).

the theory. This might lead a philosopher studying NG$_2$ to conclude that the gravitational potential should not be interpreted as a realistic feature of the world. This situation is to be contrasted with NG$_1$, where the particular value of the gravitational field at every point of spacetime, though undetermined by the empirical evidence, is nonetheless a realistic feature of the world. GNG, meanwhile, does not make any reference to a gravitational potential, and so in this sense GNG and NG$_2$ appear to have the same ontological implications, at least with regard to gravitational potentials.

However, though the gravitational potential is not invariant under the gauge transformation between models of $\mathcal{NG}_2$, some geometric properties of spacetime *are* preserved. In particular, in all models of $\mathcal{NG}_2$, *spacetime is flat*: the invariant content of the models of $\mathcal{NG}_2$ is the trajectories of bodies through a flat spacetime. In generic models of $\mathcal{GNG}$, conversely, spacetime may be curved. To put this point another way, whether or not spacetime is curved is not preserved under a natural theoretical equivalence relation. Two theories can be synonymous in the precise sense I have described, but differ with regard to their answer to what otherwise might have seemed like a clear metaphysical question. At least in this context, one can maintain that a classical spacetime admits equally good, fully equivalent descriptions as either curved or flat.

I want to emphasize that this view is not a recapitulation of Reichenbach's conventionalism. The relationship between these description is not (merely) one of empirical equivalence, and moreover, I do not think that two empirically equivalent theories need be synonymous. Rather, the point is that in the presence of certain specific structural features of Newtonian gravitational theory, properly understood, there is a stronger sense of equivalence between curved and flat descriptions of spacetime. It is not that one can freely choose between two descriptions; it is that the apparently different descriptions actually say the same thing.

This discussion suggests a more general point. Many philosophers take at least part of their work to involve identifying the metaphysical commitments of various physical theories.

But suppose one were inclined to ask what the commitments of $\mathcal{NG}_2$ should be taken to be. It might seem that $\mathcal{NG}_2$ tells us something important about the metaphysical character of spacetime—namely that it is flat. This feature, it turns out, is not preserved under a natural theoretical equivalence relation. But this raises an immediate worry: is there any way of knowing, by looking at $\mathcal{NG}_2$ in isolation, that the flatness of spacetime is not preserved under the theoretical equivalence map? It seems to me that the answer is no. And this, I think, advises caution. Given a physical theory, it often seems possible to identify what that theory says about the world, or even to use that theory to answer certain antecedent philosophical questions concerning (say) ontology or the nature of space and time. Doing so often involves identifying certain features of the theory or its models that are taken to bear on the question at hand—for instance, a philosopher who believed $\mathcal{NG}_2$ might identify the flatness of the derivative operators in all models of $\mathcal{NG}_2$ and take this to bear on a metaphysical question concerning the flatness of space and time. But what if, as in this case, the feature in question is not invariant under all theoretical equivalences?

It seems to me that in general, one cannot know in advance what features of a theory or its models will prove invariant under this theoretical equivalence map (or others), aside from the features with direct empirical consequences.[52] If this is right, then one might be skeptical of a certain kind of program in philosophy of physics, or in metaphysics, whereby one tries to read off the metaphysical commitments of contemporary physical theories, because it is not clear what features of any given theory are shared by all theories in its equivalence class. Of course, one can still learn things about the world, and about philosophical questions in particular, through the careful study of scientific theories. In the present case, for instance, the claim that there is a sense of theoretical equivalence between a description of the world in which spacetime is flat and a description of the world in which spacetime is curved provides

[52] There is something else that is preserved under the present theoretical equivalence relation, namely what one might call "the possibility structure" of the theory, in the sense of (1) what constitutes a difference in physical configuration, and (2) what the possible configurations are.

deep insight into old philosophical questions. But it seems that *what* one learns, at least in many cases, may not be quite what one had hoped. For instance, one might discover that a distinction that seemed to make perfect sense in the abstract, does not in fact correspond to a meaningful distinction within the context of a particular theory (or rather, equivalence class of theories).

By now, some readers might be worried. Indeed, one might be inclined to reject the criterion of equivalence I have proposed on the grounds that she has antecedent or even *a priori* reason for thinking that there *is* a meaningful distinction between a theory that says spacetime is flat and one that says spacetime is curved—or more generally, that she already knows what the metaphysical distinctions are, and a notion of theoretical equivalence that does not preserve the *metaphysical* structure of a theory is an unsatisfactory criterion of equivalence. The upshot would be an argument that GNG and NG must be inequivalent, since they have different metaphysical commitments.

I think this position is probably tenable, though it seems to me that it gets things backwards. At the very least, let me simply say again that there is another way of looking at matters, whereby one begins by assuming that what kinds of distinctions one can draw depend on your physical theory. On this view, what the theoretical equivalence of $\mathcal{NG}_2$ and $\mathcal{GNG}$ shows is that, within *this* theoretical context, a distinction that one otherwise thought was meaningful—i.e., the distinction between whether spacetime is curved or flat—turns out to be dependent on one's choice of (fully equivalent) representation. This point might be made more clear by pointing out that in the context of relativity theory, it would seem, one does *not* have an equivalent theory to general relativity according to which spacetime is always flat.[53] And so, I do not mean to say that one can *never* make a meaningful

[53] At least, I do not know of one. Some physicists have argued that so-called teleparallel theories of gravity offer an example of a spacetime theory in which spacetime is always flat, but which is at least empirically equivalent to general relativity (see especially Knox, 2011, and references therein). It is an interesting open question whether there is a precise sense in which teleparallel theory and general relativity are theoretically

distinction between whether spacetime is curved or flat; rather, it is that one cannot make that distinction within the framework of Newtonian gravitation (suitably understood).

Acknowledgments

Thank you to Hans Halvorson, David Malament, Jeff Barrett, John Manchak, John Norton, Kyle Stanford, Cailin O'Connor, and Noel Swanson for helpful conversations on the topics discussed here. I am particularly grateful to David Malament for detailed comments on a previous draft of this paper.

References

Andréka, H., Madaász, J. X., Németi, I., 2005. Mutual definability does not imply definitional equivalence, a simple example. Mathematical Logic Quarterly 51 (6), 591–597.

Awodey, S., 2006. Category Theory. Oxford University Press, New York.

Belot, G., 1998. Understanding electromagnetism. British Journal for the Philosophy of Science 49 (4), 531–555.

Belot, G., 2012. Symmetry and equivalence. In: Batterman, R. (Ed.), The Oxford Handbook of Philosophy of Physics. Oxford University Press, New York, forthcoming.

Borceux, F., 2008. Handbook of Categorical Algebra. Vol. 1. Cambridge University Press, New York.

Burgess, J. P., 1977. Descriptive set theory and infinitary languages. Zbornik Radova 2 (10), 9–30.

Cartan, E., 1923. Sur les variétés à connexion affine, et la théorie de la relativité généralisée (première partie). Annales scientifiques de l'École Normale Supérieure 40, 325–412.

Cartan, E., 1924. Sur les variétés à connexion affine, et la théorie de la relativité généralisée (première partie) (suite). Annales scientifiques de l'École Normale Supérieure 41, 1–25.

de Bouvere, K., 1965a. Logical synonymity. Indagationes mathematicae 27, 622–629.

de Bouvere, K., 1965b. Synonymous theories. In: Addison, J. W., Henkin, L., Tarski, A. (Eds.), The theory of models. North-Holland Pub. Co., Amsterdam, pp. 402–406.

Ehlers, J., 1981. über den Newtonschen grenzwert der Einsteinschen gravitationstheorie. In: Nitsch, J., Pfarr, J., Stachow, E.-W. (Eds.), Grundlagen Probleme der Modernen Physik. Bibliographisches Institut, Zurich, pp. 65–84.

Friedrichs, K. O., 1927. Eine invariante formulierun des newtonschen gravitationsgesetzes und der grenzüberganges vom einsteinschen zum newtonschen gesetz. Mathematische Annalen 98, 566–575.

Glymour, C., 1970. Theoretical equivalence and theoretical realism. PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association 1970, 275–288.

Glymour, C., 1977. The epistemology of geometry. Noûs 11 (3), 227 – 251.

Glymour, C., 1980. Theory and Evidence. Princeton University Press, Princeton, NJ.

Halvorson, H., 2012. What scientific theories are not, forthcoming from *Philosophy of Science*.

Halvorson, H., Swanson, N., 2012. Natural structure on state space, unpublished.

Halvorson, H., Weatherall, J. O., 2012. What is a scientific theory?, unpublished.

equivalent, or even whether they are theoretically equivalent by the standard I have proposed here. In any case, even if teleparallel gravity and general relativity *are* equivalent, the sense in which spacetime can be interpreted as flat in teleparallel theory is tendentious.

Healey, R., 2007. Gauging What's Real: The Conceptual Foundations of Gauge Theories. Oxford University Press, New York.

Hodges, W., 1993. Model Theory. Cambridge University Press, New York.

Knox, E., 2011. Newton-Cartan theory and teleparallel gravity: The force of a formulation, forthcoming from *Studies in History and Philosophy of Modern Physics*.

Künzle, H. P., 1976. Covariant Newtonian limit of Lorentz space-times. General Relativity and Gravitation 7 (5), 445–457.

Mac Lane, S., 1998. Categories for the Working Mathematician, 2nd Edition. Springer, New York.

Makowsky, J. A., Shelah, S., 1979. Theorems of beth and craig in abstract model theory. i the abstract setting. Transactions of the American Mathematical Society 256, 215–239.

Malament, D., 1986a. Gravity and spatial geometry. In: Marcus, R. B., Dorn, G., Weingartner, P. (Eds.), Logic, Methodology and Philosophy of Science. Vol. VII. Elsevier Science Publishers, New York, pp. 405–411.

Malament, D., 1986b. Newtonian gravity, limits, and the geometry of space. In: Colodny, R. (Ed.), From Quarks to Quasars. University of Pittsburgh Press, Pittsburgh, pp. 181–201.

Malament, D., 1995. Is Newtonian cosmology really inconsistent? Philosophy of Science 62 (4), 489–510.

Malament, D. B., 2012. Topics in the Foundations of General Relativity and Newtonian Gravitation Theory. University of Chicago Press, Chicago, forthcoming.

Norton, J., 1992. A paradox in Newtonian gravitation theory. PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association 1992, 412–420.

Norton, J., 1995. The force of newtonian cosmology: Acceleration is relative. Philosophy of Science 62 (4), 511–522.

Paulos, J., 1976. Noncharacterizability of the syntax set. Journal of Symbolic Logic 41 (2), 368–372.

Penrose, R., Rindler, W., 1984. Spinors and space-time. Cambridge University Press, New York.

Sklar, L., 1977. Space, Time, and Spacetime. University of California Press, Berkeley.

Sklar, L., 1982. Saving the noumena. Philosophical Topics 13, 49–72.

Spirtes, P., Glymour, C., 1982. Space-time and synonymy. Philosophy of Science 49 (3), 463–477.

Tarski, A., Givant, S. R., 1987. A formalization of set theory without variables. American Mathematical Society, Providence, RI.

Trautman, A., 1965. Foundations and current problem of general relativity. In: Deser, S., Ford, K. W. (Eds.), Lectures on General Relativity. Prentice-Hall, Englewood Cliffs, NJ, pp. 1–248.

van Fraassen, B., 2008. Scientific Representation: Paradoxes of Perspective. Oxford University Press, New York.

Weatherall, J. O., 2011. On (some) explanations in physics. Philosophy of Science 78 (3), 421–447.

Zaret, D., 1980. A limited conventionalist critique of newtonion space-time. Philosophy of Science 47 (3), 474–494.