# Respecting Boundaries: Theoretical Equivalence and Structure Beyond Dynamics

William J. Wolf

*University of Oxford*
*Oxford, OX2 6GG, United Kingdom*

James Read

*University of Oxford*
*Oxford, OX2 6GG, United Kingdom*

Abstract. A standard line in the contemporary philosophical literature has it that physical theories are equivalent only when they agree on their empirical content, where this empirical content is often understood as being encoded in the equations of motion of those theories. In this article, we question whether it is indeed the case that the empirical content of a theory is exhausted by its equations of motion, showing that (for example) considerations of boundary conditions play a key role in the empirical equivalence (or otherwise) of theories. Having argued for this, we show that philosophical claims made by Knox (2011) that general relativity is equivalent to teleparallel gravity, and by Weatherall (2016) that electromagnetism in the Faraday tensor formalism is equivalent to electromagnetism in the vector potential formalism, can both be called into question. We then show that properly considering the role of boundary conditions in theory structure can potentially restore these claims of equivalence and close with some remarks on the pragmatics of adjudications on theory identity.

Contents



*E-mail addresses*: william.wolf@philosophy.ox.ac.uk, james.read@philosophy.ox.ac.uk.
*Date*: Draft of February 14, 2023.

## 1. Introduction

Determining when two theories, models, or formulations of a theory are equivalent to one another (and in what sense) remains a significant topic within the philosophy of science (Glymour 1970; Quine 1975; Weatherall 2018). The rationale underlying the attention which has been afforded to this issue presumably has to do with the idea that it is only through understanding these issues of equivalence that one can come to understand how a theory, model, or formulation comes to limn reality. Arguably, the quest for such understanding has also aided scientific progress in the past—examples include the equivalence of wave mechanics and matrix mechanics (von Neumann 2018; Muller 1997a,b), Feynmann's and Swinger's approaches to quantum field theory (Dyson 1949), Lagrangian and Hamiltonian mechanics (Barrett 2019; Curiel 2014; North 2009), the AdS/CFT correspondence (Maldacena 1998; de Haro et al. 2016), and many others.

The extant literature on theoretical equivalence is vast, and has focused on developing criteria for—and assessing the conditions under which—particular theories can be understood as being equivalent, as well as applying these criteria to specific examples in order to illuminate our understanding of particular theories and the interconnections that their structures may possess. In a recent discussion concerning the equivalence of Lagrangian and Hamiltonian mechanics, Barrett (2019) sketches an interesting connection between questions of theoretical equivalence and questions concerning the content or structure of a physical theory. While theoretical equivalence is certainly a significant topic within the philosophy of science, Van Fraassen (1986) famously considers the question, 'what is the content of a theory?', to be the central foundational question of philosophy of science. In identifying this relationship between questions of theoretical equivalence and the content of a theory, Barrett (2019) argues that whenever we commit to a method of identifying the content of a theory, we also necessarily commit to a standard of equivalence between theories. The converse also applies because when we commit to a particular standard of equivalence between theories, we are (for Barrett) also saying which features of our theories are significant or 'contentful', as these are the very features that our assessment of equivalence will consider.

Within the physical sciences, this often manifests itself as philosophers taking the relevant physical content of a theory to be the dynamical content encoded in its equations of motion. This then necessarily results in empirical equivalence being identified with dynamical equivalence. There are a number of examples in the equivalence literature where this relationship between standards of equivalence and judgments about the content of a theory is evident. Examples include Knox (2011) arguing for the theoretical equivalence of the teleparallel equivalent of general relativity (TPG) and general relativity (GR) and Weatherall (2016) arguing for the theoretical equivalence of the Faraday tensor formulation and the gauge field formulation of classical electromagnetism. In both cases, the authors adopt a standard of equivalence that holds that the 'contentful' features of the theories in question are wholly captured in the theories' dynamics. For example, when discussing empirical equivalence, Weatherall stipulates "that on both formulations, the empirical content of a model is *exhausted* [our emphasis] by its associated Faraday tensor [that satisfies Maxwell's equations]" (Weatherall 2016, p. 1078), where of course Maxwell's equations represent the dynamical content of electromagnetism. Similarly, Knox indicates that "the (TPG) Lagrangian above turns out to be identical, up to a divergence, to the Einstein–Hilbert Lagrangian in standard GR ... the equivalence of the Lagrangians is enough to establish empirical equivalence" (Knox 2011, p. 272). This is likewise just a statement that the two theories share the same dynamics and that these dynamics exhaust empirical content. As we shall see, while they do not *explicitly* advocate a particular view of theory structure in their analyses, this standard of empirical equivalence (i.e., equivalent dynamics) nonetheless implicates both of the above authors (amongst others) in a certain fairly typical version of the semantic view of scientific theories. This view, as usually articulated, holds that a theory's content is captured by models comprised of the right kinds of mathematical objects, where these objects obey some specified dynamics.

While this is certainly an understandable position given the prominence of dynamics in physical theories, recent decades have seen both philosophers and physicists investigating content that is not entirely determined by a theory's dynamics—in particular, the content inherent to describing isolated subsystems and their relationship to their environments. Recently, philosophers have both used isolated subsystems to investigate the empirical significance of gauge symmetries (Wallace and Greaves 2014; Teh 2016; Murgueitio Ramírez and Teh forthcoming; Gomes 2021; Wolf et al. 2023), and have explored the important explanatory role that the boundary conditions associated with such isolated subsystems play in mathematical modeling (Bursten 2021). Physicists have likewise focused on isolated subsystems and boundary phenomena associated with them, as can be seen by prominent examples including the quantum Hall effect (Wen 1995), the study of black hole entropy (Gibbons and Hawking 1977), and the AdS/CFT correspondence (Maldacena 1998).

Furthermore, when viewing the content of a physical theory as including the kinds of boundary content associated with isolated subsystems, it becomes clear that an analysis of empirical equivalence that relies only on dynamics is deficient. In particular, in this paper we highlight how the analysis in the aforementioned examples from Knox and Weatherall does not account for such boundary phenomena; doing so leads to a verdict that both pairs of theories, as presented by the authors, are in fact empirically (*a fortiori* theoretically) *inequivalent*. These results thereby invite the following conclusions:

(1) Adjudications of theoretical equivalence cannot be made independently of clearly committing oneself to particular judgments regarding a theory's relevant content. If one fails to account properly for the contentful features of a

theory, one will either be left with an adjudication of theoretical equivalence that is incorrect or a view of the theories' structure that is deficient.

(2) The content of physical theories can extend beyond dynamics. Boundary phenomena, boundary conditions, and the modeling of subsystem-environment decompositions is relevant in questions concerning views of the content of physical theories, and likewise concerning theoretical equivalence, because these items are important to capturing the empirical content of physical theories. Indeed, some philosophers have begun to discuss boundary conditions alongside other elements that are typically invoked when specifying theoretical structure—see e.g. (Wallace and Greaves 2014; Teh 2016).

## 2. Views on Theoretical Equivalence

Discussions of theoretical equivalence almost invariably begin with a notion of *empirical equivalence*. If two theories disagree in terms of the empirical content associated with them, then no further analysis is necessary: they are inequivalent *tout court*. The reason for this is that empirical goings-on are naturally regarded as supervening on physical goings-on. At a minimum, theories should necessarily have the same empirical content if they are to be considered equivalent. This means that two theories must have the same range of applicability regarding empirical scenarios they describe and provide indistinguishable predictions for the observational phenomena. To be slightly more specific, we can understand that models $M$ of a theory $T$ will have empirical substructures, which can represent observable phenomena. Suppose, for every $M$ of $T$, there is an $M'$ of $T'$, where the empirical substructures of $M$ and $M'$ are isomorphic. Then, $T$ and $T'$ can be understood to be empirically equivalent (Van Fraassen 1980). This is a fairly general way of stating what empirical equivalence amounts to. As we have seen above, showing that two theories possess equivalent dynamical content is often taken to be sufficient to demonstrate empirical equivalence within the physical sciences. While we do not attempt to provide a fully exhaustive and all-encompassing definition of empirical equivalence (we can be pragmatic about this!—see §4.3), one of the goals of this paper is to demonstrate clearly that within the physical sciences there is important content *beyond* dynamics that should factor into our analyses of empirical equivalence. That is, dynamical equivalence alone is not sufficient to establish empirical equivalence.

Those of a positivist persuasion would consider empirical equivalence to be a sufficient criterion for establishing theoretical equivalence because they would argue that a theory's meaning and content is exhausted by its empirical consequences. Yet, most subscribe to the idea that empirical equivalence is a necessary but not sufficient condition for theoretical equivalence, because there are meaningful theoretical claims beyond strict empirical consequences, such as two theories differing in regards to "what structure they attribute to the world, what sorts of entities exist in the world, or what the laws of nature are" (Weatherall 2018, p. 5). This has motivated philosophers to propose further, stronger criteria for establishing theoretical equivalence that go beyond empirical consequences. These can be roughly broken down into formal notions of equivalence and interpretational equivalence. This literature is vast and we make no attempt at a fully exhaustive description of the possibilities.

*Definitional equivalence* is a formal criterion developed initially by the likes of Quine (1975) and Glymour (1970, 1980), and captures the idea that two theories should be inter-translatable. This means that one should be able to take all of the vocabulary of theory $T$, and translate it into the vocabulary of theory $T'$, and *vice versa*, in a manner that faithfully preserves the content of each theory. Furthermore, there is generally an idea that these translations between theories should be unique and invertible. Other formal attempts at cashing out equivalence in something like this manner include categorical equivalence and Morita equivalence. *Categorical equivalence* uses tools from category theory to address situations that seem otherwise to be problem cases for definitional equivalence, such as when transformations between models are many-to-one (Weatherall 2018). This is the case, for example, when multiple gauge choices in one theory correspond to one model on the other side of the transformation. *Morita equivalence* attempts to weaken definitional equivalence by providing a notion of equivalence that applies to theories that are formulated using different sorts (i.e., different classes of entities) (Barrett and Halvorson 2016).

*Interpretational equivalence*, in contrast with definitional equivalence, seeks to capture the notion that two theories are equivalent when they license all of the same claims about the phenomena they describe, going beyond purely empirical or formal considerations (Coffey 2014; Butterfield forthcoming). In other words, theories $T$ and $T'$ can be understood to postulate the same ontologies and make the same claims about this shared ontology.

With these notions of equivalence on the table, we next move on to analyzing some recent discussions in the philosophical literature surrounding the issue of theoretical equivalence, and to evaluating these respective adjudications of theoretical equivalence for particular theories. The examples we will consider are (i) the equivalence of the teleparallel equivalent of general relativity (TPG) and general relativity (GR) (Knox 2011), and (ii) the equivalence of the Faraday tensor formulation and the gauge field formulation of electromagnetism (Weatherall 2016).

## 3. Adjudicating Theoretical Equivalence

3.1. **Example 1: TPG and GR.** Both TPG and GR are theories of gravitation, but TPG differs from general relativity in a number of ways. The most obvious is that rather than using the symmetric Levi-Civita connection $\Gamma^{\rho}_{\mu\nu}$, the unique torsion-free connection, TPG uses the Weitzenböck connection $\dot{\Gamma}^{\rho}_{\mu\nu}$, which has a non-vanishing torsion and vanishing curvature. That is, rather than expressing gravity as a manifestation of spacetime curvature as GR does, TPG holds that gravity is a manifestation of spacetime torsion. TPG views gravity as a force because torsion directs bodies experiencing gravitation away from geodesics, as opposed to the situation in GR, whereby bodies experiencing gravitation follow the geodesics resulting from spacetime curvature. Furthermore, TPG is usually formulated in terms of tetrads $e^a_\mu$, rather than a metric $g_{\mu\nu}$. Tetrads, or frame fields, are sets of four linearly independent fields $e^a = e^a_\mu dx^\mu$ that at each point $p$ of a differentiable manifold $M$ specify a basis for the tangent space $T_pM$.[1] TPG uses frame fields $h^a_\mu = e^a_\mu + B^a_\mu$ that are constructed to be invariant under local translations $x^a \mapsto x^a + \epsilon^a$, where $B^a_\mu$ is the translation gauge potential. This gauge potential transforms as $\delta B^a_\mu = -\partial_\mu \epsilon^a$ so as to make the frame field invariant under such local translations. It is for this reason that TPG is often declared to be a gauge theory of the translation group (Aldrovandi and Pereira 2013).[2]

TPG and GR are seemingly very different theories, constructed using different mathematical structures—but Knox (2011) has argued that GR and TPG should in fact be understood as being equivalent to one another. She argues for this conclusion based upon: (i) the establishment of dynamical equivalence (and thus, for her argument, empirical equivalence) and definitional equivalence between the two theories, and (ii) an interpretation of TPG that holds that both TPG and GR actually postulate the same underlying spacetime structure despite the surface level appearances, which appears to be motivated by her advocacy of spacetime functionalism. As before, theoretical equivalence is taken to be a combination of demonstrating empirical equivalence, along with stronger notions of equivalence that demonstrate clear formal relations between the theories and resolve interpretive issues such that we can understand both theories as making the same claims about the target phenomena. While the spacetime functionalism component of her argument certainly brings up a host of interesting issues, this is not the place to fully adjudicate the interpretational issues she raises regarding TPG and GR. However, we would like to focus specifically on the discussion of empirical equivalence between the theories.

The claim that TPG and GR are empirically equivalent is motivated by appealing to actions used in each theory,

$$S_{TPG} = \frac{1}{16\pi G}\int d^4x\sqrt{h}T, \qquad S_{GR} = \frac{1}{16\pi G}\int d^4x\sqrt{g}R, \tag{1}$$

where $h$ is the determinant of the tetrad, $T$ is the torsion scalar defined as $T = \mathcal{S}_\rho^{\ \mu\nu} T^\rho_{\ \mu\nu}$, $\mathcal{S}_\rho^{\ \mu\nu}$ is the so-called superpotential tensor, $T^\rho_{\ \mu\nu}$ is the torsion tensor, $g$ is the determinant of the metric, and $R$ is the Ricci scalar. The superpotential tensor is built out of the torsion tensor and the so-called contorsion tensor. The contorsion tensor is defined as $K^\rho_{\ \mu\nu} := \Gamma^\rho_{\ \mu\nu} - \dot{\Gamma}^\rho_{\ \mu\nu}$, where we see that it is simply the difference between the Weitzenbock connection, $\dot{\Gamma}^\rho_{\ \mu\nu}$, and the Levi-Civita connection, $\Gamma^\rho_{\ \mu\nu}$.

[1] What we have in fact written here are cotetrad fields $e^a$, which are those 1-forms such that $e^a_\mu e^\mu_b = \delta^a_b$; we do so since this will simplify the presentation in what follows.

[2] For some critical discussion of this claim, see (Wallace 2015).

This is significant because this allows one to translate between the mathematical structures of the teleparallel theory and those of general relativity. One can use this to re-write the TPG action in the language of GR as[3]

$$S_{TPG} = \frac{1}{16\pi G}\int d^4x\sqrt{g}R + \frac{1}{8\pi G}\int d^4x\sqrt{g}\nabla_\mu T^{\alpha\mu}_\alpha. \tag{2}$$

This shows that the TPG action is identical to the Einstein-Hilbert action of GR plus a total divergence term, which ensures that these actions both lead to the same dynamical equations of motion. On the basis of these observations, Knox makes three arguments regarding the equivalence of GR and TPG:

**Empirical equivalence:** The equivalence of the actions up to a total divergence term, which indicates that they both share equivalent dynamics, guarantees the empirical equivalence of TPG and GR (Knox 2011, p. 272).

**Definitional equivalence:** The relationship between the Levi-Civita connection and the Weizenböck connection allows us to directly translate between GR and TPG and vice-versa. While definitional equivalence is not explicitly mentioned in her argument, this is a clear appeal to a similar notion of equivalence. Anything we express in the language of GR can be equivalently expressed in the language of TPG and *vice versa* in a way that preserves the content of each theory. For example, we have already seen how one moves between different connection coefficients and translates between spacetime curvature and torsion, but one can similarly translate between the frame fields of TPG and the metric of GR as $g_{\mu\nu} = \eta_{ab}h^a_\mu h^b_\nu$, where $\eta_{ab}$ is the Minkowski metric (Knox 2011, p. 272).

**Interpretational equivalence:** TPG and GR both encode the same spacetime structure, upon adopting spacetime functionalism (which, for Knox, is the view that spacetime structure is whatever identifies a class of local inertial frames—for critical discussion of this view, see e.g. (Read and Menon 2021)), and thus can be understood as licensing the same claims about the phenomena they describe (Knox 2011, p. 273).

The argument that the actions are empirically equivalent hinges on the ability to throw away the total divergence term present in (2). Once this term is discarded, the actions are equivalent full stop and the argument for definitional equivalence goes through as well because these terms can be safely ignored when making these kinds of translations between TPG and GR. But why can this total divergence term simply be thrown away?

When discussing a particular theory whose content is captured by an action $S$, typically one takes the empirical content of that theory to be derived via a variational principle.[4] The 'principle of least action' is a variational principle which holds

[3] see (Aldrovandi and Pereira 2013, ch. 9) for a fully explicit derivation.

[4] This isn't to say that variational principles *exhaust* the empirical content of theories—for example, conserved quantities can be derived via Noether's theorems. These points will not matter for our purposes here.

that the variation of the action is held fixed when the equations of motion—i.e., the dynamics—of the system are satisfied. Consider the simple textbook example of a free massive particle in motion where our variables are position $q(t)$ and velocity $\dot{q}(t)$ and the action is given by $S = \int_{t_i}^{t_f} L[q, \dot{q}, t]dt$:

$$\delta S = \int_{t_i}^{t_f} \left[\frac{\partial L}{\partial q} - \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}}\right)\right] \delta q dt + \frac{\partial L}{\partial \dot{q}} (t_f)\, \delta q\, (t_f) - \frac{\partial L}{\partial \dot{q}} (t_i)\, \delta q\, (t_i) = 0. \tag{3}$$

Here we find the familiar Euler-Lagrange equations of motion in the first term. However, we also have two further terms which are the result of a total divergence that appears after the integration by parts necessary to write the Euler-Lagrange equations in their standard form. In this case we are simply concerned with the motion of a particle between two fixed end points, $\delta q(t_i)$ an $\delta q(t_f)$. These remaining terms thus automatically go to zero, leaving just the dynamics of our system captured in the first term. These total divergence terms do not affect the underlying dynamics of the system; furthermore, it is important to emphasize that any terms like this *must* vanish for there to be a well-defined variational principle at all, as a proper functional derivative could not be defined otherwise.

Given that we typically throw away total divergence terms because we know that they have to vanish anyway, our work is seemingly done. The TPG action encodes the same dynamics as the GR action, so the equations of motion will be the same and we are left to choose the language in which to express them: the force equations of TPG or the geodesic equations of GR. That is, $\delta S_{GR} = \delta S_{TPG} = \frac{1}{16\pi G} \int d^4x \sqrt{-g} G_{\mu\nu} \delta g^{\mu\nu}$, where $G_{\mu\nu}$ contains the dynamical equations of motion. Thus, "the equivalence of the Lagrangians is enough to establish empirical equivalence" (Knox 2011, p. 272). The question of the theoretical equivalence between TPG and GR then hinges only upon the interpretive questions.

When doing GR, we typically consider only manifolds without boundary. This guarantees that the total divergence term in (2) is zero because Stokes' theorem allows us to convert a total divergence term into a boundary term. In the event that there is no boundary, this term automatically vanishes. For example, this is exactly what is done in using GR to model cosmological solutions as we are attempting to describe the entire universe and its contents filling an infinite space. What if we wanted to model some isolated subsystem instead? Consider an isolated subsystem $\mathcal{S}$ that is being modeled with respect to an external environment $\mathcal{E}$. For example, we might be interested in determining the mass-energy content of a finite region of spacetime, such as the mass-energy content contained within a black hole, as defined by an external observer who is sufficiently far away so that they do not interact with any of the relevant gravitational or material fields. In this event, it is not appropriate to consider manifolds without boundary. Rather, the manifold $M$ must have a boundary $\partial M$ along with appropriate boundary conditions to properly describe a subsystem $\mathcal{S}$ isolated from its environment $\mathcal{E}$. Total divergence terms such as the one we have considered then cannot be automatically discarded and generally will not vanish.

When considering the Einstein-Hilbert action in the presence of the boundary $\partial M$, such residual total divergence terms are indeed present and we must find appropriate boundary conditions to render this a well-defined variation.[5] Here, it is natural to consider Dirichlet boundary conditions, $\delta g_{\mu\nu}|_{\partial M} = 0$, as these boundary conditions correspond to asymptotically flat spacetimes. These are spacetimes that approach flatness $g_{\mu\nu} \to \eta_{\mu\nu}$ at null-infinity and are particularly significant for a number of reasons. Here is Penrose (1982) on the issue:

> Asymptotically flat spacetimes are interesting, not because they are thought to be realistic models for the entire universe, but because they describe the gravitational fields of isolated systems, and because it is only with asymptotic flatness that general relativity begins to relate in a clear way to many of the important aspects of the rest of physics, such as energy, momentum, radiation, etc.

That is, in the asymptotic regime we can clearly define critical, empirically relevant concepts such as mass, energy, and momentum, and relate them to these concepts as they are understood in other realms of physics. (In brief: in the asymptotic regime, one has Killing fields, with which one can associate conserved quantities in a well-understood way: see e.g. (De Haro 2021).)

Upon imposing Dirichlet boundary conditions $\delta g_{\mu\nu}|_{\partial M} = 0$, we find that there is a problem. There are multiple boundary terms and it is only the term that depends on the tangential derivatives of the metric that vanishes, while another term that depends on the normal derivatives survives.[6] This is because Dirichlet boundary conditions fix only the values of the metric of the boundary, but this does not necessarily require the values of derivatives of the metric to vanish. In other words, the variation of this action does not yield a well-defined variation and cannot be used to represent or model isolated subsystems of the type that Penrose refers to in his description of asymptotically flat spacetimes. This is closely related to what

[5]Recalling the Einstein-Hilbert action of GR, considering a manifold with a boundary $\partial M$, and varying the action yields

$$\delta S_{\mathrm{GR}} = \frac{1}{16\pi G}\int_M d^4x\sqrt{-g}G_{\mu\nu}\delta g^{\mu\nu} + \frac{1}{16\pi G}\oint_{\partial M} d^3\Omega\sqrt{h}\left(g^{\sigma\nu}\delta\Gamma^{\rho}_{\nu\sigma} - g^{\sigma\rho}\delta\Gamma^{\mu}_{\mu\sigma}\right), \tag{4}$$

where $G_{\mu\nu}$ is the Einstein tensor, $g_{\mu\nu}$ is the metric tensor, $h_{\mu\nu}$ is the induced metric on the boundary, and $\Gamma^{\mu}_{\nu\lambda}$ represents the connection coefficients of the Levi-Civita connection (Blau n.d., ch. 20.2). The first term yields via variation the Einstein field equations and vanishes when the dynamics of the theory are satisfied. The remaining term comes from a total divergence that has been converted to a boundary term via Stokes' theorem.

[6]It is helpful to rewrite the boundary term from the GR action eq. (4) as:

$$\frac{1}{16\pi G}\oint_{\partial M} d^3\Omega\sqrt{h}\left(N^{\rho}h^{\mu\nu}\nabla_{\mu}\delta g_{\rho\nu} - N^{\mu}h^{\rho\nu}\nabla_{\mu}\delta g_{\rho\nu}\right), \tag{5}$$

where $N$ is the vector normal to the boundary (Blau n.d., ch. 20.5). The first term depends on tangential derivatives of the metric and its variation, whereas the second term depends on normal derivatives of the metric and its variation. Dirichlet boundary conditions kill only the first term, while leaving the second term intact.

Belot (2018) observes when he emphasizes that two isomorphic solutions in GR do not always represent the same physical possibilities. In Belot's analysis, he notes that cosmological solutions and asymptotically flat solutions are isomorphic dynamically, but do not represent the same physical possibilities. This is because the boundary conditions imposed for each solution are physically relevant facts. This discussion of the variational problem in GR reveals that the Einstein-Hilbert action only has the resources to represent one of these two physical possibilities (cosmological solutions), and that we need to look elsewhere to represent asymptotically flat solutions. We see that even within GR, dynamically equivalent solutions do not necessarily represent the same physical possibility. Thus, merely demonstrating the dynamical equivalence between a GR action and a TPG action likewise would not necessarily indicate that the two theories are physically equivalent.

This indicates the importance of boundary conditions in specifying the content of our theory and the scope of the empirical scenarios and target systems that our models and theories can represent. Let us now compare the analogous scenario in TPG to see how the teleparallel theory fares in describing isolated subsystems with aysmptotic characteristics.

Amazingly, upon varying the TPG action and imposing Dirichlet boundary conditions, we find that the TPG action indeed does have a well-defined variation (Oshita and Wu 2017). The variation of the additional boundary term that distinguishes the TPG and Einstein-Hilbert actions ensures that the total variation is well-defined for asymptotic spacetimes because the additional terms perfectly cancel out the previously problematic terms.[7] The reason for this can be traced to the fact that the TPG action contains only first derivatives of the frame fields, whereas the Einstein-Hilbert formulation contains second derivatives of the metric. The additional boundary term effectively removes the second derivatives of the metric that fail to vanish when working with the Einstien-Hilbert action.

This TPG action functions perfectly well for describing such isolated subsystems. For anyone who may understandably be perturbed by the thought that GR cannot describe such systems: do not worry. The ways in which GR handles these situations will be important later. The point is that as this argument for theoretical equivalence is presently formulated (in terms of these two dynamically equivalent actions), TPG and GR are *not* empirically equivalent—and so, as per the above,

[7] Recall that TPG differs from the GR action by a total divergence term. Therefore, we can apply Stokes' theorem to the divergence term in eq. (2), and add this to the result above. Converting the boundary term in eq. (4) to the language of TPG frame fields, adding the additional TPG boundary term, and imposing Dirichlet boundary conditions $\delta g_{\mu\nu}|_{\partial M} = 0$, $\delta e^a_\mu\big|_{\partial M} = 0$ yields the following (Oshita and Wu 2017):

$$\begin{aligned}\delta S_{\mathrm{TPG}} = &\frac{1}{16\pi G}\int_M d^4x\sqrt{-g}G_{\mu\nu}\delta g^{\mu\nu} + \frac{\epsilon}{8\pi G}\oint_{\partial M} d^3\Omega\sqrt{h}n^\mu\left[e^\alpha_A\partial_\alpha\delta e^A_\mu - e^\alpha_A\partial_\mu\delta e^A_\alpha\right]\\ &+ \frac{\epsilon}{8\pi G}\oint_{\partial M} d^3\Omega\sqrt{h}n^\mu\left[e^\alpha_A\partial_\mu\delta e^A_\alpha - e^\alpha_A\partial_\alpha\delta e^A_\mu\right]\\ = &\, 0.\end{aligned}$$

should not be regarded as being equivalent, full stop. Under this articulation, these theories do not even have the resources to describe all of the same target systems, much less discuss whether one can compare the empirical consequences derived from them for said target systems.

3.2. **Example 2: Faraday Tensor EM and Gauge Field EM.** This is not the only example from the recent philosophical literature on equivalence where there has been a proclaimed theoretical equivalence between two theories that relies on understanding empirical equivalence as dynamical equivalence. Weatherall (2016) examines two different formulations of classical electromagnetism which are typically taken to be theoretically equivalent to each other. These two formulations are $\mathrm{EM}_1$, where electromagnetism is presented in terms of the Faraday tensor $F_{\mu\nu}$, and $\mathrm{EM}_2$, where electromagnetism is presented in terms of the electromagnetic potential field $A_\mu$. This example is slightly different from the previous example of TPG and GR (where there is some serious uncertainty regarding equivalence) as here there is a near-universal consensus amongst both physicists and philosophers that $\mathrm{EM}_1$ and $\mathrm{EM}_2$ are in fact theoretically (hence empirically) equivalent. Here the issue is that actually cashing out this consensus with a philosophical model of equivalence is a slightly thornier business than one would initially expect due to the gauge freedom present in the $\mathrm{EM}_2$ formulation.

Weatherall proposes that while these two different formulations do not meet the standard criteria for definitional equivalence as proposed by Glymour (1970, 1980) due to there being non-isomorphic translations between formulations, one can use categorical equivalence to capture the theoretical equivalence of these formulations. However, his argument for full theoretical equivalence also depends on establishing that models derived from these different formulations really do capture *all* of the same empirical content and are thus empirically equivalent as well. We find that the manner in which he argues for empirical equivalence follows a similar pattern to Knox in that there is important empirical content that is missing from his account due to the overly restrictive view that empirical content is exhausted by dynamical content. To stress (and to repeat), there is an overwhelming consensus from both physicists and philosophers that $\mathrm{EM}_1$ and $\mathrm{EM}_2$ *are* empirically equivalent. We are not challenging the notion that they are in fact equivalent, but rather challenging the philosophical criteria used in this analysis because they fail to capture the empirical equivalence that these two formulations readily display within the practice of physics.

Examining Weatherall's analysis more closely, we find that he utilizes a conceptual framework that is broadly consistent with the semantic conception of scientific theories. We will have more to say on this later, but essentially the standard articulation of the semantic view holds that theories are collections of dynamically equivalent models. Weatherall takes $\mathrm{EM}_1$ to be a theory given by models built out of the following objects: $\langle M, \eta_{\mu\nu}, F_{\mu\nu} \rangle$, where $M$ is a smooth manifold, $\eta_{\mu\nu}$ is the

Minkowski metric, and $F_{\mu\nu}$ is the Faraday tensor. These models furthermore must all satisfy the dynamics encoded by Maxwell's equations (1) $\nabla_{[\rho}F_{\mu\nu]} = 0$ and (2) $\nabla_\mu F^{\mu\nu} = J^\nu$, where $J^\nu$ is the charge density current. $EM_2$ is a theory given by $\langle M, \eta_{\mu\nu}, A_\mu \rangle$, where $A_\mu = (\phi, \vec{A})$ is the four-potential vector field and these models likewise satisfy Maxwell's equations in the form $\Box A^\mu = J^\mu$, where $\Box$ is the D'Alembertian operator.[8] Weatherall's analysis quite understandably holds that these two 'theories' or 'formulations' of a single theory (whichever you prefer), are empirically equivalent:

> **Empirical equivalence:** "We stipulate that on both formulations, the empirical content of a model is exhausted by its associated Faraday tensor. In this sense, the theories are empirically equivalent, since for any model of $EM_1$, there is a corresponding model of $EM_2$ with the same empirical content (for some fixed $J^a$), and vice versa" (Weatherall 2016, p. 1078). In other words, $EM_1$ and $EM_2$ both share all of the same dynamical content and are thus empirically equivalent.[9]

As we all know, these different formulations are very closely related. Given the Faraday tensor $F_{\mu\nu}$ that satisfies Maxwell's equations, there is always a vector field $A_\mu$ that also satisfies Maxwell's equations and satisfies the definition $F_{\mu\nu} = \nabla_{[\mu}A_{\nu]}$. Similarly, given a vector field $A_\mu$ that satisfies Maxwell's equations, there is always a corresponding tensor $F_{\mu\nu}$ that satisfies Maxwell's equations and can be defined as $F_{\mu\nu} = \nabla_{[\mu}A_{\nu]}$ (all of these facts follow from elementary properties of differential forms). As Weatherall notes, however, one cannot find an isomorphism between the spaces of models of these two formulations of classical electromagnetism. Starting with the $EM_2$ formulation, given a vector potential $A_\mu$, one can uniquely define a Faraday tensor $F_{\mu\nu}$ in $EM_1$. Conversely, going in the other direction and given a Faraday tensor $F_{\mu\nu}$ in $EM_1$, one cannot uniquely determine a model in $EM_2$ due to the gauge freedom present in the four-potential $A_\mu$. That is, $F_{\mu\nu}$ is compatible with infinitely many different $A_\mu$ because $F_{\mu\nu} = \nabla_{[\mu}A_{\nu]}$ will hold for any $A_\mu$ such that $A'_\mu = A_\mu + G_\mu$ if $\nabla_{[\mu}G_{\nu]} = 0$, or in other words, if $G_\mu$ is a closed one-form. Given that a straightforward application of definitional equivalence is blocked, Weatherall motivates modifying Glymour's criterion for definitional equivalence (i.e., for every model in $T$, there is an isomorphic translation to a model in $T'$ that preserves all of the same empirical content) in terms of isomorphisms between categories of models that preserve empirical content. Here, we now understand the models of $EM_2$ to be $\langle M, \eta_{\mu\nu}, [A_\mu] \rangle$, where here $[A_\mu]$ is an "equivalence class of physically equivalent vector potentials" that correspond to the same $F_{\mu\nu}$ (Weatherall 2016, p.1079). Note that this adjustment depends on the argument that $EM_1$ and $EM_2$

[8] When we refer to 'Maxwell's equations' in what follows, strictly speaking we equivocate between the formulation of these equations in terms of $F_{\mu\nu}$ and the formulation in terms of $A_\mu$; that said, which version we have in mind should always be evident from context.

[9] One might be puzzled here by Weatherall's lack of mention of the Aharonov-Bohm effect on this point; we will not dwell on this issue further in this article.

are actually empirically equivalent, and this equivalence class of vector potentials essentially is constructed to wash out the gauge parameter so that the translations between formulations are isomorphisms.

**Categorical equivalence:** Categorical equivalence is stated in terms of categories of models that preserve empirical content. Thus, according to Weatherall's construction, we can uniquely and invertibly translate between models of $\text{EM}_1$ and $\text{EM}_2$ and their respective vocabularies, provided that $\text{EM}_2$ is redefined such that $[A_\mu]$ is an equivalence class of vector potentials that lead to the same $F_{\mu\nu}$. Then, we have $\langle M, \eta_{\mu\nu}, F_{\mu\nu} \rangle \iff \langle M, \eta_{\mu\nu}, [A_\mu] \rangle$, where "there exists an isomorphism between their categories of models that preserves empirical content" (Weatherall 2016, pp. 1080-1) and this further notion of formal equivalence is then used to argue that both formulations are theoretically equivalent.

While it is certainly correct that all models in both formulations possess the same dynamical content, as we have seen that does not mean that they necessarily share all of the same empirical content. After all, we just saw in the previous section how multiple actions can share the same dynamical content, while still differing in the totality of their empirical content due to their differing capacities to represent certain target phenomena. As before, the idea that dynamics capture the full empirical content of a theory is present in this argument for theoretical equivalence. Yet again, when considering boundaries, we will find that there is important and relevant empirical content that goes beyond dynamics.

The following example is significantly more familiar than what we considered in the previous section on TPG and GR. Consider an environment-subsystem decomposition that includes a simple Faraday cage, described by a finite spatial subsystem region with a surface boundary $\partial M$. This subsystem region has boundary conditions conducive to describing a perfect electric conductor with a surface charge $\sigma$ and this Faraday cage serves to shield the interior of the subsystem from electromagnetic fields in the environment. Let us now consider $\text{EM}_1$ and $\text{EM}_2$ models of the Faraday cage.

Beginning with the Faraday tensor formulation $\text{EM}_1$, this construction in terms of the electric and magnetic fields will lead to the conclusion that the Faraday tensor describing the subsystem is *always* zero. This is simply a consequence of the fact that regardless of what the external electric and magnetic fields are, the conducting boundary will always arrange the surface charge $\sigma$ to cancel the effect of the external fields. Thus, $F_{\mu\nu} = 0$ inside the cage regardless of facts about the external fields and surface charge. By contrast, the gauge field formulation $\text{EM}_2$ shows that the gauge potentials describing the subsystem will instead be constant. While so far this is all consistent as this is what we expect of potentials that lead to $F_{\mu\nu} = 0$, as Murgueitio Ramírez and Teh (forthcoming) emphasize in their paper concerning the direct empirical significance of gauge symmetries, specifying the scalar electric potential $\phi$ on the boundary uniquely specifies the surface charge $\sigma$ on the boundary. Furthermore, in general one can fully construct a solution for $\phi$ for both the subsystem and exterior in terms of the surface charge $\sigma$, which comes back to the fact that shifts in the potential shift the charge at the boundary surface.[10]

How do these considerations influence our verdict on the empirical equivalence of these two formulations? $EM_1$ treats the Faraday tensor as the fundamental object of interest. The same Faraday tensor $F_{\mu\nu}$ within the isolated subsystem could potentially correspond to two empirically distinct surface charges $\sigma_1$ and $\sigma_2$ (in fact, it corresponds to infinitely many different surface charges!). However, $EM_2$ treats gauge potentials as the fundamental objects of interest and the gauge potential $\phi$ always distinguishes between $\sigma_1$ and $\sigma_2$. To be completely explicit, let us adopt Weatherall's initial characterization of $EM_1$ and $EM_2$ as models given by $\langle M, \eta_{\mu\nu}, F_{\mu\nu}\rangle$ and $\langle M, \eta_{\mu\nu}, A_\mu\rangle$, respectively.[11] Furthermore, let us say that we are interested in an empirical description of a Faraday cage with a surface charge $\sigma_1$. On Weatherall's characterization, $EM_1$ corresponds to $\langle M, \eta_{\mu\nu}, 0\rangle$ and $EM_2$ corresponds to $\langle M, \eta_{\mu\nu}, \phi_i(\sigma_1)\rangle$, where $\phi_i$ is the scalar potential for the subsystem. This $EM_1$ description could correspond to infinitely many subsystems all with different surface charges because they will all lead to $F_{\mu\nu} = 0$, whereas the $EM_2$ description uniquely describes the subsystem with the particular surface charge we are considering here. In other words, the model of the subsystem in $EM_2$ has the information necessary to model empirical facts about boundary phenomena and the external environment, information that the model of the subsystem in $EM_1$ simply does not have when we hold that the structural content of a theory is given exclusively in terms of mathematical objects and dynamics. Essentially, the mathematical objects that the respective formulations are built out of carry different amounts of empirical information. On this reading, these descriptions of the subsystem are not empirically equivalent because they do not carry the same empirical information about the target system and you cannot deduce the same empirical consequences from them.

It is important to emphasize here that we are not arguing against the empirical equivalence of the Faraday tensor and gauge field formulations of electromagnetism. In other words, a physicist can deduce the same empirical claims about

[10]This happens in three steps, following Murgueitio Ramírez and Teh (forthcoming). (1) One introduces a gauge transformation $A \to A' + d\chi(x,t)$, where $\chi$ is the gauge parameter. (2) One fixes the gauge parameter by choosing the Coulomb gauge by $\nabla \cdot A' = \nabla \cdot (A + \nabla\chi(x,t)) = 0$. (3) This then leads to Poisson's equation for the scalar electric potential $\phi(x)$. They note that this procedure reveals interesting features of the gauge parameter $\chi(x,t)$. That is, $\chi(x,t)$ is a field-dependent parameter that depends on the four-potential $A$ and in gauge-fixing this parameter, one shifts the electric scalar potential $\phi(x) \mapsto \phi'(x) = \phi(x) - \partial_t\chi(x,t)$ both in the interior and boundary of the subsystem. To make this more explicit, $\phi'$ satisfies the Poisson equation, where at the boundary $\nabla^2\phi' = \sigma$. We can then recognize a relationship between $\chi$ and $\sigma$ through $\nabla^2(\phi(x) - \partial_t\chi(x,t)) = \sigma$. As specifying the potential $\phi$ uniquely specifies the surface charge $\sigma$, shifts in the gauge parameter $\chi$ also induce shifts in the surface charge $\sigma$.

[11]Here we are concerned with empirical equivalence which is independent of the categorical equivalence issue that Weatherall addresses in his later characterization of $EM_2$

such a subsystem from both formulations as using these formulations in practice involves specifying further items (like the boundary conditions and their relationship to surface charges) that are necessary to build the electromagnetic fields and potentials relevant to describing the system. Rather, we are arguing that the philosophical criteria for evaluating empirical equivalence in terms of dynamics alone is insufficient to account for the equivalence of $\mathrm{EM}_1$ and $\mathrm{EM}_2$. Indeed, this view leads to the shocking conclusion that $\mathrm{EM}_1$ and $\mathrm{EM}_2$ (when stated as $\langle M, \eta_{\mu\nu}, F_{\mu\nu}\rangle$ and $\langle M, \eta_{\mu\nu}, A_{\mu}\rangle$ respectively) are not equivalent because there is significantly more empirical information contained within a model specified as $\langle M, \eta_{\mu\nu}, A_{\mu}\rangle$. As before, the account of empirical equivalence gets tripped up when considering isolated subsystems and external environments. However, as we shall soon see, a more nuanced view of theory structure and empirical equivalence can restore our intuition that both formulations of electromagnetism are equivalent to each other.

## 4. Views on Theory Structure

What is happening here? We have two fairly prominent examples of arguments for the theoretical equivalence of the respective theories considered in these examples. One of these examples (TPG and GR) is more contentious given the extent of the interpretive arguments that need to be made to secure interpretational equivalence, but the other (Faraday tensor and vector potential formulations of EM) is utterly uncontroversial. Yet, as articulated, these arguments for theoretical equivalence cannot even support claims of empirical equivalence for these respective theories. Something has clearly gone wrong!

Perhaps it is the way in which the theories have been stated that has disrupted these claims of empirical equivalence. After all, in making an adjudication of theoretical equivalence, it is certainly important to correctly specify the empirical content contained by a theory. Views on the structure of scientific theories can be roughly broken down into three camps: the ‘syntactic’, ‘semantic’, and ‘pragmatic’ views. The syntactic view seeks to axiomatize a theory in terms of abstract mathematical sentences. The semantic view casts a theory in terms of models and the kinds of mathematical objects that comprise these models. While the syntactic view was initially dominant as it emerged first as an outgrowth from logical empiricism, van Fraassen has prominently advocated for the semantic view by arguing that the semantic view, with its focus on models, can often more simply demonstrate the logical claims of a theory than a set of axioms.[12] Furthermore, he argues that the semantic view is a far more comprehensive and useful tool because it avoids the restrictions inherent to describing a theory in a particular axiomatic language, and

[12]Indeed, van Fraassen acknowledges that one can often derive the same logical claims concerning the statements a theory makes about the world from both approaches, with the caveat that these claims are more clearly and simply expressed on the semantic view. Lutz (2017) has taken this further and argued that both syntactic and semantic views are actually far more closely related than has been supposed in the literature and the debate surrounding which approach is preferable is largely illusory.

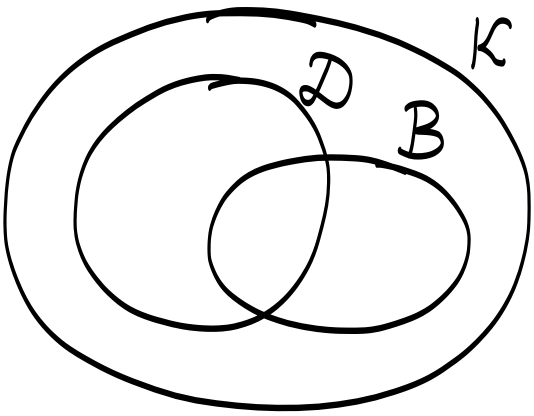
allows us to conceptualize the objects and classes of structures that comprise a model in terms of a variety of valid, non-unique descriptions (Van Fraassen 1980, pp. 43–4). Finally, the pragmatic view is a more recent perspective that emphasizes representational aims, model pluralism, scientific practice, and other non-formal characteristics (Cartwright 1983; Hacking 1983; Kitcher 1993; Winther 2021). In this article, we will focus on viewing these adjudications of theoretical equivalence through the lens of both the semantic and pragmatic views.

4.1. **The Semantic View.** The semantic view of theories holds that a theory is individuated via class of models. One modern way of expressing the semantic view is to say that a theory $\mathcal{T}$ has a set of 'kinematically possible models' $\mathcal{K}$ (KPMs), defined by tuples of the form $\langle O_i, ...O_n \rangle$, where these $O_i$ are mathematical objects, e.g. tensor fields on a differentiable manifold. Furthermore, these objects come with a set of particular dynamical equations that define the relationships and interactions between the $O_i$. KPMs that satisfy these dynamical equations form a subspace $\mathcal{D} \subset \mathcal{K}$ of KPMs known as the 'dynamically possible models' (DPMs). In other words, "the KPMs can be thought of as representing the range of metaphysical possibilities consistent with the theory's basic ontological assumptions. The DPMs represent a narrower set of physical possibilities" (Pooley 2013, p. 12). This dynamical content is then understood to capture the empirical content of the models that comprise the theory, via what van Fraassen calls the 'empirical substructures' of each of these models (Van Fraassen 1980, p. 45).

It is clear that Weatherall draws from this framework in his analysis. For example, his descriptions of $\text{EM}_1$ and $\text{EM}_2$ as theories with associated respective classes of models $\langle M, \eta_{\mu\nu}, F_{\mu\nu} \rangle$ and $\langle M, \eta_{\mu\nu}, A_\mu \rangle$ identifies the relevant KPMs, where his specification that these models obey Maxwell's equations identifies the particular DPMs that correspond to the theories in question.

While the utilization of the standard semantic view is not as obvious in Knox, it is clear that something like this is being supposed in her identifying the theory of GR with the empirical content contained within the Einstein-Hilbert action. Recall that in her argument it is the local equivalence of the two actions that cements the case for empirical equivalence, which really is just the statement that both theories share the dynamical content when the actions are varied per standard variational principles. In identifying the Einstein-Hilbert action as capturing GR's content and adjudicating the empirical equivalence of GR and TPG based on the dynamical equivalence of these actions, there is a naturally consistency with the standard semantic expression of GR in the philosophical literature.

In more detail: in the above-introduced model-based language (Pooley 2013, 2015), GR is usually given by KPMs of the form $\langle M, g_{\mu\nu}, \Phi \rangle$, where (again) $M$ is a smooth, four dimensional differentiable manifold, $g_{\mu\nu}$ is the metric tensor field on $M$, and $\Phi$ represents the matter fields of the theory. The DPMs of GR are the subset of the KPMs that obey the Einstein equation, which is given by $G_{\mu\nu} = 8\pi T_{\mu\nu}$,




where $G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu}$ is the familiar Einstein tensor and $T_{\mu\nu}$ is the stress-energy tensor. For Knox, the Einstein-Hilbert action contains all of these objects we are interested in that comprise the kinematic possibilities of GR, and varying this action gives us the dynamics that these models obey. We could likewise identify TPG with the KPMs $\langle M, e^a_\mu, \Phi \rangle$, whose DPMs are the subset of KPMs that also obey the Einstein field equations (written in terms of the primitive objects of TPG, i.e. the objects specified in the KPMs of that theory).

Read through this lens, both Weatherall and Knox are operating within a framework whereby they are identifying the relevant empirical content of the theories they are interested in with the dynamics obeyed by the models that comprise these theories. It is a very straightforward argument. There are theories given by models of the form $\langle M, g_{\mu\nu}, \Phi\rangle$ and $\langle M, e^a_\mu, \Phi \rangle$, as well as $\langle M, \eta_{\mu\nu}, F_{\mu\nu}\rangle$ and $\langle M, \eta_{\mu\nu}, A_\mu\rangle$. The first pair obeys the dynamics encoded by the Einstein field equations and the second pair obeys the dynamics encoded by the Maxwell equations. Therefore, both pairs are empirically equivalent to each other. The key assumption, of course, is that dynamics fully specifies the empirical content of these theories and the models that comprise them. Yet, as we have already seen, there is important empirical content that this characterization leaves out.

4.2. **Boundary Possible Models.** The above discussion invites a modification of the now-standard KPM/DPM version of the semantic approach. Here, we introduce a third class of models—proposed by Read (2016)—known as 'boundary possible models' $\mathcal{B}$ (BPMs). Here, $\mathcal{B} \subset \mathcal{K}$, and would denote the subset of KPMs compatible with particular boundary conditions. Then those $\mathcal{B} \cup \mathcal{D} \subset \mathcal{K}$ would specify those KPMs that are compatible with both particular boundary conditions and particular dynamics.

Coming back to the example of TPG, the action $S_{TPG}$ and its variation $\delta S_{TPG} = 0$ captures the empirical content for models that are compatible with both Dirichlet boundary conditions and the Einstein field equations. That is $S_{TPG}$ gives us the subset of the KPMs that satisfies Dirichlet boundary conditions and the dynamics of the Einstein field equations $\mathcal{B}_D \cup \mathcal{D}_{EFE}$. As we saw, while the Einstein-Hilbert action (which we will now switch to specifying as $S_{EH}$) shares the same dynamics $\mathcal{D}_{EFE}$, it is not capable of representing isolated subsystems with the Dirichlet boundary conditions $\mathcal{B}_D$. This invites the question: can the models derived from the Einstein-Hilbert action represent any isolated subsystems and can isolated subsystems with Dirichlet boundary conditions be modeled within the framework of GR at all?

The answer to the former question is that there are boundary conditions that make the Einstein-Hilbert action well-defined. Recall that the normal derivatives of the metric did not vanish when examining the boundary terms in eq. (5). Neumann boundary conditions, rather than specifying the values of the metric on the boundary, specify the values of the metric's derivatives on the boundary. It turns out that when one imposes suitable Neumann boundary conditions, both the terms involving tangential and normal derivatives with respect to the metric vanish (Freidel et al. 2021). We can then clearly see with this framework that $S_{EH}$ and $S_{TPG}$ do not share the same empirical content because $\mathcal{B}_N \cup \mathcal{D}_{EFE} \neq \mathcal{B}_D \cup \mathcal{D}_{EFE}$ (recall again that $S_{TPG}$ gives us the subset of the KPMs $\mathcal{B}_D \cup \mathcal{D}_{EFE}$).

Finally, how does GR actually model isolated subsystems with Dirichlet boundary conditions and study important concepts found in asymptotic spacetimes? The answer is that we must set aside the Einstein-Hilbert action $S_{EH}$ in favor of what is known as the Gibbons-Hawking-York (GHY) action $S_{GHY}$:

$$S_{\mathrm{GHY}} = \frac{1}{16\pi G}\int d^4x\sqrt{-g}R + \frac{1}{8\pi G}\oint_{\partial M} d^3\Omega\epsilon\sqrt{h}K, \tag{6}$$

where $K = \nabla^\mu n_\mu$ is the trace of the extrinsic curvature, $h$ is the induced metric on the boundary, and $\epsilon$ is $+1$ when the boundary hypersurface is spacelike and $-1$ when the boundary hypersurface is timelike (York 1972; Gibbons and Hawking 1977). We see here that this action is equal to the Einstein-Hilbert action plus a boundary term. When varying this action, we find the bulk term that contains the dynamical Einstein field equations $G_{\mu\nu}$, the boundary term from before, and a further boundary term originating from the GHY term. Upon imposing Dirichlet boundary conditions $\delta g_{\mu\nu}|_{\partial M} = 0$, we find that the variation of the GHY boundary term exactly cancels out the previously non-vanishing terms. Thus, in the presence of manifolds with boundaries with Dirichlet boundary conditions, we have $\delta S_{\mathrm{GHY}} = \frac{1}{16\pi G}\int_M d^4x\sqrt{-g}G_{\mu\nu}\delta g^{\mu\nu}$. This follows the exact same pattern as the variation of the TPG action. The additional boundary term plays a similar role and cancels out previously problematic terms, yielding a well-defined variation.

We see that $S_{GHY}$ gives us the subset of KPMs $\mathcal{B}_D \cup \mathcal{D}_{EFE}$. This matches up with the subset of KPMs given to us by $S_{TPG}$, which as we have seen is also $\mathcal{B}_D \cup \mathcal{D}_{EFE}$. Indeed, $\delta S_{TPG} = \delta S_{GHY}$ when $\mathcal{B}_D$ is imposed, so we know that both theories share the same dynamical content and the same representational capacity when it comes to isolated subsystems. Important quantities that depend on these boundary terms and conditions such as the ADM mass $M_{ADM}$ and black hole entropy $S_{BH}$ are found to be in agreement. For example, $M_{ADM}$ is one of the quantities to which

Penrose referred and represents the mass-energy content contained within a finite region of spacetime. Using $S_{GHY}$ and $S_{TPG}$ to determine this quantity gives the same results and these results are crucially dependent on the role and behavior of the boundary terms and conditions that we have discussed (Dyer and Hinterbichler 2009; Wald 1993; Iyer and Wald 1994; Hammad et al. 2019). As Freidel and Teh have noted, these boundary terms can also effectively bring the Noether charges of a theory into alignment with the corresponding Hamiltonian charges (i.e., the ADM mass), which connects such quantities to Hamiltonian observables (Freidel and Teh 2021). Coming to black hole entropy $S_{BH}$, one can use the Euclidean semi-classical path integral approach and find that one obtains identical results for this quantity and the boundary terms present in both $S_{TPG}$ and $S_{GHY}$ contribute the entire entropy in the calculation (Gibbons and Hawking 1977; Gibbons et al. 1978; Oshita and Wu 2017).

Our conception of a theory should specify the empirical content of the theory. KPMs define the objects of interest to us within a particular theory, but we would not say that defining a theory exclusively in terms of KPMs is satisfying because it plainly fails to do this. We also want to specify how these objects interact with each other and behave empirically. DPMs specify their dynamics. However, as the above examples demonstrate, dynamics does not constitute the full extent of the empirical content of these models. We also want to specify the subsystem-environment decompositions that these models can represent, as well as any boundary related empirical content that goes beyond the dynamics of these objects. Just as KPMs are insufficient to fully specify a theory's empirical content, nor are DPMs alone: the latter should be supplemented with BPMs to more fully specify to empirical content of a theory.

Coming back to the issues of empirical and theoretical equivalence, it is clear that one's conception of a theory will have a non-trivial impact on any subsequent adjudication of theoretical equivalence. The identification of GR with the dynamics resulting from the Einstein-Hilbert action and of $\text{EM}_1$ with a Faraday tensor obeying Maxwell's equations does not fully specify the empirical content of those theories and thus is responsible for incorrect adjudications of empirical equivalence when compared with their allegedly equivalent counterparts. Both Knox and Weatherall do make some qualifying statements. Knox notes that the local equivalence of the TPG and EH actions up to a divergence may lead to some global worries (Knox 2011, p. 272), while Weatherall notes that he assumes that the empirical content of EM is exhausted by Faraday tensors compatible with Maxwell's equations (Weatherall 2016, p. 1078). Yet, it is clear that in both cases, there are indeed global worries that render their adjudications problematic and that these qualifying statements do not do justice to the empirical content that is lost when one looks exclusively at local dynamics.

How could one go about arguing for theoretical equivalence of GR and TPG given our characterization of the semantic view that includes KPMs, DPMs, and

BPMs? One way would involve taking inspiration from the characterization of equivalence found in Nguyen (2017). This would mean showing that models from both GR and TPG can represent the same target systems, and that they make the same empirical claims about these target systems. We have already partially done that by showing $S_{GHY}$ and $S_{TPG}$ coincide in the target systems they represent and discussing how they align in the empirical claims they make about boundary dependent phenomena that goes beyond the shared dynamics of all these models. One could similarly investigate other actions, models, and isolated subsystems in both GR and TPG and ensure that they align in both representational capacity and empirical claims. This still leaves open the admittedly more difficult interpretative questions regarding whether GR and TPG license all of the same interpretive claims about the world and their target systems, but it at least provides a straightforward path to perspicuously demonstrating their empirical equivalence.

For $EM_1$ and $EM_2$, the key is simply to realize that boundary conditions are essential information in any attempt to represent a subsystem-environment decomposition. That is, whether we are using $EM_1$ or $EM_2$, we must specify boundary conditions in order to actually build solutions for the mathematical objects that those descriptions make use of (electric and magnetic fields versus gauge fields, respectively). Yet, under the standard semantic way of expressing these theories (in terms of mathematical objects and dynamics), for a system like the Faraday cage one formulation contains more empirical information than the other precisely because the boundary conditions we used in building the mathematical objects are left out of the formal description of the theory. If boundary conditions are admitted to the formal criteria that defines the structure of a theory, this incongruity dissolves because these boundary conditions contain the information that is needed for the Faraday tensor formulation $EM_1$ to distinguish between different surface charges from within the Faraday cage; ie, something that $EM_2$ naturally does because the surface charges find their way into the gauge potentials. Thus, it then becomes clear that $EM_1$ and $EM_2$ are indeed empirically equivalent once we admit boundary conditions into the semantic criteria.

4.3. **The Pragmatic View.** We can also draw from the pragmatic view of theories to illuminate these adjudications of theoretical equivalence and the importance of having a firm view of the structure of the theories in question. Rather than totally repudiating the syntactic and semantic views, the pragmatic view acknowledges the utility of many of the formal components of these other perspectives, while also emphasizing non-formal considerations. While there is significant variety amongst proponents of this view (Cartwright 1983; Hacking 1983; Kitcher 1993; Winther 2021), two strands of thought stand out as particularly relevant to the present discussion: (i) model pluralism and (ii) focus on scientific practice.

On (i): Cartwright claims that models are the appropriate level of scientific investigations (as opposed to theories), and that "models serve a variety of purposes, and individual models are to be judged according to how well they serve the purpose at hand" and notes that there are many different but legitimate reasons to utilize different models (Cartwright 1983, p. 152). One model might be focused on accuracy for a particular quantity, while another might be trying to incorporate additional phenomena into the description and consequently, might be less focused on maximizing accuracy of any one particular quantity.

This point is made quite generally, but we can see something similar going on in GR. We have already encountered two actions used in GR, $S_{EH}$ and $S_{GHY}$, but there are others, including but not limited to the Gamma-Gamma action $S_{\Gamma\Gamma} = \frac{1}{16\pi G}\int d^4x\sqrt{-g}g^{\mu\nu}\left(\Gamma^{\alpha}_{\mu\beta}\Gamma^{\beta}_{\alpha\nu} - \Gamma^{\alpha}_{\mu\nu}\Gamma^{\beta}_{\alpha\beta}\right)$ and the ADM action $S_{ADM} = \frac{1}{16\pi G}\int d^4x\sqrt{-g}(\tilde{R} + K^{\mu\nu}K_{\mu\nu} - K^2)$, where $\tilde{R}$ is the three-dimensional Ricci scalar of the spatial slice in the in the $3+1$ decomposition in the $ADM$ formulation and $K$ is the extrinsic curvature. $S_{\Gamma\Gamma}$ turns out to be incredibly convenient for demonstrating that GR corresponds to the self-coupling of a massless spin-2 particle, due to the cubic nature of the form of the Lagrangian, which is in analogy with both Yang-Mills fields and spin-1 particles and chiral fields and spin-zero particles (Deser 1970, 1987). Additionally, $S_{ADM}$ is particularly important because it leads to a Hamiltonian interpretation of GR that, among many other benefits, has become crucial in numerical relativity due to Hamiltonian systems leading to first-order equations of motion rather than the second-order equations of motion one derives from Lagrangians. This has opened the door for GR to study and model far more rich and complex physical possibilities than could ever have been done using pure analytic methods (Pretorius 2005), including the modeling of binary black hole mergers that proved crucial in the eventual detection of gravitational waves (Abbott et al. 2016). Whether we choose an action based on convenience, clarity, or necessity, there are a lot of options at our disposal for modeling phenomena in GR. Under this pragmatic approach of embracing model pluralism, it is clear that GR is much broader than the dynamical content of one of these actions and any adjudication of theoretical equivalence would need to address this broader scope.

Another theme that the pragmatic view emphasizes—point (ii) above—is that our view of theories should be commensurate with scientific practice. While acknowledging the utility of formal criteria, Teh has argued that a theory should be more properly viewed as a collection of physical representations, "accompanied by a keen 'know how' about what we can do with such representations and how they are related to each other" (Teh forthcoming, p. 7). This emphasis on 'know how' implores us to consider scientific practice in specifying the structure of theories and has indeed been a major focus of advocates for the pragmatic view (Hacking 1983; Kitcher 1993). Clearly, practitioners of GR use many different dynamically equivalent actions depending on the problem at hand, but this discussion also highlights how boundary phenomena has become more relevant in both physics and philosophy communities in more recent years. As we have already noted physics itself has been exploring boundary phenomena and isolated subsystems in recent decades with examples including the quantum Hall effect, black hole entropy, and AdS/CFT correspondence, while philosophy has been interested in them as a way to cash out the direct empirical significance of symmetries and the explanatory capabilities of models.

Here, we see a potential connection between the semantic and pragmatic approaches. While the pragmatic view does emphasize non-formal elements of modeling and theory structure, its embrace of pluralism also allows it to accommodate a variety of strategies in describing theory structure, including the use of more formal notions. Indeed, some philosophers have even argued that "the semantic conception in its bare minimal expression" is very compatible with "pragmatic elements and themes" (Suárez and Pero 2019, p. 348). We can thus rely on pragmatic considerations such as scientific practice to inform us of what structures should find their way into a formal representations of the models in our theories. Before the theoretical and empirical importance of boundaries was truly appreciated, it might have made more sense to view a theory exclusively in terms of its dynamics and mathematical objects. However, as scientific practice (and philosophical interest) has changed and brought this boundary phenomena more into focus, it now makes sense to adjust our views on the structure of theories to be commensurate with scientific practice. As we saw in the previous section, one can easily accommodate boundary conditions within a tradition semantic analysis of a theory.

## 5. Consequences and Conclusions

Discussions concerning both the equivalence and structure of physical theories have been and will continue to be important themes in the philosophy of science. As we have seen (following Barrett (2019)), each of these questions bears upon the other because adopting a particular standard of equivalence will necessarily specify a view of what the contentful features of a theory actually are; and similarly, adopting a particular view of theory content or structure will necessarily set the standard by which equivalence is to be judged.

The aforementioned examples in the literature regarding the supposed theoretical equivalence between TPG and GR and between $EM_1$ and $EM_2$ illustrate both that these questions do indeed interact with each other and suggest that these questions need to be tackled in parallel. In navigating these issues surrounding theory equivalence and structure, we take one moral from this discussion to be that adopting a pragmatic attitude towards theory structure can be very fruitful. Indeed, we saw that in both examples considered, the source of the failure of empirical equivalence came about from the authors adopting a view of theory structure that, while useful and fairly standard throughout the philosophy literature, used formal criteria that were overly restrictive regarding the empirical substructures that one could attribute to the theories. Thus, in adopting this standard semantic view, additional empirical content related to boundary phenomena and isolated subsystems did not make its way into the analysis.

Recalling Knox's analysis of TPG and GR, this creates a scenario where adopting the standard semantic view indicates that both theories are empirically equivalent, but this equivalence is hollow because it is based off a deficient view of the theories' structure and empirical content, as the importance and richness of boundary phenomena has come into full view for both physicists and philosophers alike. In other words, this equivalence only goes through when one omits certain empirical content. When we properly consider these further empirical substructures stemming from boundary phenomena, the argument for empirical equivalence—as articulated based upon analyzing the dynamical content derived from two particular actions—fails because the theories (again, as articulated) are clearly not empirically equivalent due to their inability to represent all of the same target systems.

However, the pragmatic view can help bring these discussions of equivalence and structure into alignment. As we have seen in these examples, the pragmatic view indicates that we should be pluralistic regarding the actions we use to model phenomena in these theories and allows us to pragmatically update the components we consider when utilizing formal descriptions of theory structure by supplementing the standard semantic representation with boundary conditions. In so doing, we can construct an argument for the equivalence of TPG and GR that also reflects the full richness of the empirical content that these theories are currently understood to possess. While the example of $EM_1$ and $EM_2$ is not quite as dramatic given that there are not multiple actions to choose from in representing these formulations of the theory, there is something similar going on. When boundary conditions are included in the formal criteria that describe theory structure, it is clear that $EM_1$ and $EM_2$ are equivalent as well and that the issue merely stemmed from adopting an overly restrictive view of the empirical content contained within the formal descriptions. Furthermore, this pragmatic attitude provides flexibility in that it allows us to continuously update our understanding of theory structure as previous empirical substructures become better understood and novel empirical substructures come into view. In the context of these many empirical realizations surrounding boundary phenomena and their increased importance to both physicists and philosophers, it is clear that such an update is needed and that boundary conditions and phenomena must be considered in discussions of the empirical content that a theory is understood to possess.

## Acknowledgements

We are grateful to Caspar Jacobs, Ruward Mulder, and Nic Teh for helpful discussions. This work was supported by grant number 61521 from the John Templeton Foundation.

## References

Abbott, B. P. et al. (2016). Observation of Gravitational Waves from a Binary Black Hole Merger, *Phys. Rev. Lett.* **116**(6): 061102.

Aldrovandi, R. and Pereira, J. G. (2013). *Teleparallel Gravity: An Introduction*, Springer.

Barrett, T. W. (2019). Equivalent and inequivalent formulations of classical mechanics, *British Journal for the Philosophy of Science* **70**(4): 1167–1199.

Barrett, T. W. and Halvorson, H. (2016). Morita equivalence, *Review of Symbolic Logic* **9**(3): 556–582.

Belot, G. (2018). Fifty million elvis fans can?t be wrong, *Noûs* pp. 946–981.

Blau, M. (n.d.). Lecture notes on general relativity.
**URL:** *http://www.blau.itp.unibe.ch/newlecturesGR.pdf*

Bursten, J. R. S. (2021). The function of boundary conditions in the physical sciences, *Philosophy of Science* **88**(2): 234–257.

Butterfield, J. (forthcoming). On dualities and equivalences between physical theories, *in* C. Wüthrich, B. L. Bihan and N. Huggett (eds), *Philosophy Beyond Spacetime*, Oxford: Oxford University Press.

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

Coffey, K. (2014). Theoretical equivalence as interpretative equivalence, *The British Journal for the Philosophy of Science* **65**(4): 821–844.
**URL:** *https://doi.org/10.1093/bjps/axt034*

Curiel, E. (2014). Classical mechanics is lagrangian; it is not hamiltonian, *The British Journal for the Philosophy of Science* **65**(2): 269–321.

De Haro, S. (2021). Noether's Theorems and Energy in General Relativity.

de Haro, S., Mayerson, D. and Butterfield, J. (2016). Conceptual aspects of gauge/gravity duality, *Foundations of Physics* **46**(11): 1381–1425.

Deser, S. (1970). Selfinteraction and gauge invariance, *Gen. Rel. Grav.* **1**: 9–18.

Deser, S. (1987). Gravity from self-interaction in a curved background, *Classical and Quantum Gravity* **4**(4): L99–L105.
**URL:** *https://doi.org/10.1088/0264-9381/4/4/006*

Dyer, E. and Hinterbichler, K. (2009). Boundary Terms, Variational Principles and Higher Derivative Modified Gravity, *Phys. Rev. D* **79**: 024028.

Dyson, F. J. (1949). The radiation theories of tomonaga, schwinger, and feynman, *Phys. Rev.* **75**: 486–502.

Freidel, L., Oliveri, R., Pranzetti, D. and Speziale, S. (2021). Extended corner symmetry, charge bracket and Einstein's equations, *JHEP* **09**: 083.

Freidel, L. and Teh, N. (2021). Substantive general covariance and the Einstein-Klein dispute: A Noetherian approach.

Gibbons, G. W. and Hawking, S. W. (1977). Action Integrals and Partition Functions in Quantum Gravity, *Phys. Rev. D* **15**: 2752–2756.

Gibbons, G. W., Hawking, S. W. and Perry, M. J. (1978). Path Integrals and the Indefiniteness of the Gravitational Action, *Nucl. Phys. B* **138**: 141–150.

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

Glymour, C. (1980). Theory and evidence, *Ethics* **93**(3): 613–615.

Gomes, H. (2021). Holism as the empirical significance of symmetries, *European Journal for Philosophy of Science* **11**(3): 1–41.

Hacking, I. (1983). *Representing and Intervening: Introductory Topics in the Philosophy of Natural Science*, Cambridge University Press.

Hammad, F., Dijamco, D., Torres-Rivas, A. and Bérubé, D. (2019). Noether charge and black hole entropy in teleparallel gravity, *Phys. Rev. D* **100**(12): 124040.

Iyer, V. and Wald, R. M. (1994). Some properties of Noether charge and a proposal for dynamical black hole entropy, *Phys. Rev. D* **50**: 846–864.

Kitcher, P. (1993). *The Advancement of Science: Science Without Legend, Objectivity Without Illusions*, Oxford, England: Oxford University Press.

Knox, E. (2011). Newton?cartan theory and teleparallel gravity: The force of a formulation, *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics* **42**(4): 264–275.

Lutz, S. (2017). What was the syntax-semantics debate in the philosophy of science about?, *Philosophy and Phenomenological Research* **95**(2): 319–352.

Maldacena, J. M. (1998). The Large N limit of superconformal field theories and supergravity, *Adv. Theor. Math. Phys.* **2**: 231–252.

Muller, F. A. (1997a). The equivalence myth of quantum mechanics –part i, *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics* **28**(1): 35–61.

Muller, F. A. (1997b). The equivalence myth of quantum mechanics–part ii, *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics* **28**(2): 219–247.

Murgueitio Ramírez, S. and Teh, N. (forthcoming). Abandoning galileo's ship: The quest for non-relational empirical significance, *British Journal for the Philosophy of Science* .

Nguyen, J. (2017). Scientific representation and theoretical equivalence, *Philosophy of Science* **84**(5): 982–995.

North, J. (2009). The structure of physics: A case study, *Journal of Philosophy* **106**(2): 57–88.

Oshita, N. and Wu, Y.-P. (2017). Role of spacetime boundaries in Einstein's other gravity, *Phys. Rev. D* **96**(4): 044042.

Penrose, R. (1982). Some unsolved problems in classical general relativity.

Pooley, O. (2013). Substantivalist and relationalist approaches to spacetime, *in* R. Batterman (ed.), *The Oxford Handbook of Philosophy of Physics*, Oxford University Press.

Pooley, O. (2015). Background independence, diffeomorphism invariance, and the meaning of coordinates, *in* D. Lehmkuhl (ed.), *Towards a Theory of Spacetime Theories*, Birkhäuser.

Pretorius, F. (2005). Evolution of binary black-hole spacetimes, *Physical review letters* **95**(12): 121101.

Quine, W. (1975). On empirically equivalent systems of the world, *Erkenntnis* **9**(3): 313–28.

Read, J. (2016). *Background independence in classical and quantum gravity*, University of Oxford.

Read, J. and Menon, T. (2021). The limitations of inertial frame spacetime functionalism, *Synthese* .

Suárez, M. and Pero, F. (2019). The representational semantic conception, *Philosophy of Science* **86**(2): 344–365.

Teh, N. (forthcoming). *Philosophy of Symmetry*, Cambridge University Press.

Teh, N. J. (2016). Galileo?s gauge: Understanding the empirical significance of gauge symmetry, *Philosophy of Science* **83**(1): 93–118.

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

Van Fraassen, B. C. (1986). Aim and structure of scientific theories, *in* R. Barcan Marcus, G. J. Dorn and P. Weingartner (eds), *Logic, Methodology and Philosophy of Science VII*, Vol. 114 of *Studies in Logic and the Foundations of Mathematics*, Elsevier, pp. 307–318.

von Neumann, J. (2018). *Mathematical Foundations of Quantum Mechanics: New Edition*, Princeton University Press.
**URL:** *https://doi.org/10.1515/9781400889921*

Wald, R. M. (1993). Black hole entropy is the Noether charge, *Phys. Rev. D* **48**(8): R3427–R3431.

Wallace, D. (2015). Fields as bodies: a unified presentation of spacetime and internal gauge symmetry.
**URL:** *https://arxiv.org/abs/1502.06539*

Wallace, D. and Greaves, H. (2014). Empirical consequences of symmetries, *British Journal for the Philosophy of Science* **65**(1): 59–89.

Weatherall, J. (2016). Are newtonian gravitation and geometrized newtonian gravitation theoretically equivalent?, *Erkenntnis* **81**(5): 1073–1091.

Weatherall, J. O. (2018). Theoretical equivalence in physics.

Wen, X.-G. (1995). Topological orders and edge excitations in FQH states, *Adv. Phys.* **44**(5): 405–473.

Winther, R. G. (2021). The Structure of Scientific Theories, *in* E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*, Spring 2021 edn, Metaphysics Research Lab, Stanford University.

Wolf, W. J., Read, J. and Teh, N. J. (2023). Edge Modes and Dressing Fields for the Newton–Cartan Quantum Hall Effect, *Found. Phys.* **53**(1): 3.

York, Jr., J. W. (1972). Role of conformal three geometry in the dynamics of gravitation, *Phys. Rev. Lett.* **28**: 1082–1085.