Interpreting GR as a guideline for theory change

Niels Linnemann*†

To appear in: Claus Beisbart, Tilman Sauer, and Christian Wüthrich, editors: Thinking About Space and Time: 100 Years of Applying and Interpreting General Relativity. Birkhäuser (Basel), Forthcoming.

1 Introduction

100 years of relativity — but what is next? In this essay, I explore to what extent general relativity (GR) itself suggests particular (candidate) successor theories — or rather, to what extent the various empirical and conceptual interpretational stances one can take concerning GR do. At a more general level, the essay thus aims at demonstrating how interpretational questions allow for systematically generating hypotheses about a successor theory to GR, and thus conceiving of the theory change from GR as — at least potentially — a well-guided process.

In section 2, I first clarify the notion of interpretation by using a three-fold distinction by Curiel (2009). I will be concerned with the straightforward empirical interpretation of a theory, that is what Curiel calls concrete interpretation (sometimes also just empirical interpretation in the following), and the high-level conceptual interpretation of a theory, that is what Curiel calls categorial

*University of Geneva, niels.linnemann@unige.ch

†I would like to thank Christian Wüthrich, James Read, Kian Salimkhani, Manus Visser, Patrick Dürr and Rasmus Jaksland for comments on this manuscript. I would in particular like to thank James Read for discussions on EPS and radar coordinates, which are picked up on in detail in a joint manuscript in progress. Note that this paper is an extract from my PhD thesis "Philosophy of quantum gravity as a philosophy of discovery" which was generously funded by the Swiss National Science Foundation (SNF) under the project number 105212_165702. The content of this work is solely the responsibility of the author and does not represent the official views of the SNF.

interpretation. In section 3, I consider the rather unusual question in how far empirical interpretation (concrete interpretation) can provide a guideline for a successor theory to GR. I identify two potential strategies for suggesting a direction in theory change from empirical interpretation: (1) On the one hand, we can ask at a general level what other theory-external descriptions of the established empirical content of the theory there are, and whether these should not be incorporated as well. (2) On the other hand, we can wonder whether specific options for increasing the empirical content of GR and the empirical access to it *via using external theories* can encourage the unification of GR with exactly these external theories. In section 4, I deal with the more common case of motivating successor theories to GR from its conceptual interpretations; I present some examples of how different conceptual takes on GR encourage successor theories to it.

2 Empirical interpretation(s) of a physical theory

In dealing with the question of how to make sense of GR's empirical content, Curiel (2009) distinguishes three notions of empirical interpretation:

> **Concrete.** The fixation of a semantics for the formalism, in the sense that the formalism under the semantics expresses the empirical knowledge the framework contains—for example, the fixation of a Tarskian family of models, or, less formally, the contents of a good, comprehensive textbook.
> **Categorial.** The explication of concepts in the theory that the semantics of a concrete interpretation alone does not fix—for example, a demonstration that the theory is deterministic in any of a variety of senses.
> **Metalinguistic.** The explication of the semantics of a concrete interpretation, when the representational nature of the concrete interpretation is itself not understood—for example, the Copenhagen Interpretation of quantum mechanics. (p. 46)[1]

[1] Admittedly, it is not really clear why this interpretation is best dubbed 'metalinguistic', since—as we will see—it can also be understood as a fine-grained correction or addition to the concrete interpretation.

Following Curiel, *all* physical theories require a concrete and a categorial interpretation. More precisely: (1) All physical theories have to make contact with the world in some way or the other in order to count as empirical theories. Thus, there should be no dispute about that physical theories need to be interpreted concretely. (2) Whether a categorial interpretation is necessary or not, does not amount to a case-by-case question but rather to a general question of interest independent of the nature of specific physical theories. Examples for categorial questions include whether a theory is deterministic or indeterministic, whether it allows for superluminal propagation or whether it is observer-independent or -dependent. At least from the view of a philosopher of physics, *all* physical theories should be worth being interpreted categorially in some way or another. And in any case, a categorial question is either raised for all/a subset[2] of theories or not at all. More metaphysically-flavoured questions such as whether a spacetime theory is substantivalist or relationist etc. are arguably categorial as well. We will, however, ignore questions like this in the following as we are interested in empirical senses of interpretation only.

In contrast, only a few physical theories, such as QM[3] require a metalingustic interpretation. That necessity of a metalinguistic interpretation is a case-by-case question should be clear from that Newtonian mechanics, and QM respectively provide paradigm examples for theories which do not have and have to be interpreted metalinguistically. As an illustration, consider that in both Newtonian mechanics and QM the Hamiltonian's value is assigned in one way or the other to energy. Whereas in the former, this assignment holds without restriction, this is not at all the case in QM:[4] The Hamiltonian H is assigned to a measurable

[2]Certain categorial questions such as whether a theory allows for superluminal propagation require a background framework to be even sensibly formulated in the first place. They thus only make sense for a subset of theories admissible to this overarching framework. See for instance Weatherall (2014) for a discussion of theory-overarching criteria for superluminal propagation.

[3]Note that by 'usual', we mean textbook QM. This is not to say that the concrete interpretation of QM cannot drastically vary either. GRW, for instance, is not just a metalinguistic interpretation of QM but at the same time has a different concrete interpretation than standard QM. Within GRW, the characteristic spontaneous localisation process (including an avalanche of localisations induced by a spontaneous localisation) is part of the empirically measurable content.

[4]Cf. Curiel (2009), p. 47:

> The way that Hermitian operators in standard quantum mechanics represent observables is perhaps the canonical example of such a problem: we know they do in some way or other represent observables, and we know how to use them to construct good models of systems that we can use to predict the (probabilistic) outcomes of experiments, but we have no clear understanding at all of the nature of the representational relations between, on the one hand, the operator as part of the formalism and, on the other, the actual values we measure for physical

energy E in the concrete interpretation but then further qualification is needed: The Hamiltonian H can only be assigned to a measurable energy E upon some story of collapse of a generic state into an eigenstate of H, as $H|\psi\rangle = E|\psi\rangle$ iff $|\psi\rangle$ is an eigenstate of H.

(Strictly speaking, QM only has to be given some kind of qualificatory interpretation for which a metalinguistic interpretation might be just one way to do so: The Copenhagen interpretation fleshes out in how far operators are 'observables' and in particular in what sense experiments show determinate results by providing the primitive theory-external notion of collapse — this is an instance of metalinguistic interpretation. As an alternative to this, the many-worlds interpretation arguably provides a non-trivial internal qualification of what happens in measurement through invoking the idea of a virtual, branch-relative equivalent of an objective collapse. Both the Copenhagen as well as the many-worlds 'interpretations' can be seen as qualificatory interpretations on top of the *usual* concrete interpretation of QM. For a detailed explication of the notion of qualificatory interpretation, which is beyond the scope of this short overview on notions of empirical interpretations, see Linnemann (2019).)

2.1 The concrete interpretation in more detail

The notion of concrete interpretation can be straightforwardly cashed out in the rather narrow positivist notion of partial interpretation (see Suppe (2000)):[5]

quantities in experiments.

[5] Note that a sensible partly interpretation (which the concrete interpretation amounts to) does not have to follow mere empirical interests. For instance, Lehmkuhl (2017) exemplarily pleads for a partly (what he calls 'careful') rather than literal interpretation of the Schwarzschild metric field even within a realist context when the metric is invoked in modelling the perihelion motion of Mercury:

> What Einstein really does is to convert the two-body problem Sun-Mercury into a one-body problem, where one body (Mercury) is subject to an external gravitational field. ... It is the exterior gravitational field of the Sun, not the Sun itself, that is represented by the Schwarzschild metric. And that is enough to predict the perihelion of Mercury: we do not need to know what the Sun is made of or what happens in its interior; all that matters is the exterior gravitational field that Mercury is subject to. Thus, worrying about the singularity at the center of the Schwarzschild metric misses the point: we do not have to interpret the interior part of the Schwarzschild metric literally, at least not in this application. (pp. 1210-1211)

So, the Schwarzschild metric's interior structure should not be seen as referring literally (not even approximately) but rather as a placeholder for any kind of interior structure that (1) would be compatible with the exterior metric field of the Schwarzschild solution, and (2) would involve structure which can in the end be straightforwardly interpreted as referring to the sun.

> For positivists, theories were partially interpreted axiomatic systems TC where the axioms T were the theoretical laws expressed in a theoretical vocabulary V_T; C were correspondence rules that connected T with testable consequences formulated using a separate observational vocabulary V_O. Only V_O sentences were given a direct semantic interpretation. (p. S103)

On such a syntactic understanding of theories[6], the concrete interpretation fixes (and it alone fixes) by assignments the semantics necessary to test the theory for *empirical adequateness*. It is commonly claimed that the syntactic view of theories has by now been superseded by the semantic view of theories[7] for a set of well-known reasons: among other things, the syntactic view (unlike the semantic view) is usually seen to (1) suffer from issues of theory individuation ($T_1 = p \rightarrow q$ and $T_2 = \neg p \vee q$ are logically equivalent but — oddly — different syntactic theories), (2) make an untenable theoretical/observational distinction at the level of language (rather than that of entities), and (3) narrowly require theories to be formulated in first-order logic (see Winther (2016), and Suppe (2000). Recent arguments, however, render the syntactic view as capable of dealing with these challenges (see, for instance, Lutz (2014)), and as basically equivalent to the semantic view (see Halvorson (2013) and Lutz (2017)).

Since we are only interested in denoting the empirical content of a theory, we strictly speaking better stay uncommitted to the existence of the non-observable part of the formalism (constructive empiricism). With the syntactic view, we, however, run into the so-called closure problem[8]: given that belief statements are normally closed under logical implication, fixing referents for the empirical (and just for the empirical) on the syntactic view of theories will automatically commit one to that certain theoretical terms refer as well. A statement which mixes observable and theoretical terms such as "A table is a swarm of electrons, protons and neutrons." would, for instance, ground the existence of electrons, protons and neutrons in the world based on that we believe that tables exist in the world. This would run counter the idea that the concrete interpretation, when defined syntactically as above, is *just* an empirical interpretation of the

[6]The syntactic view renders a theory as a set of statements T closed under logical implication (subset of statements, called axioms, generates T when requiring logical closure), and correspondence rules between the entities in T and the world. See Winther (2016), section 2.

[7]The semantic view renders a theory as a set of models. In the context of a physical theory, you can loosely think of each solution to the dynamical equations for a given set of initial conditions as a model. See Winther (2016), in particular section 3.1.2, and references within. A locus classicus is Van Fraassen (1980).

[8]First introduced by Friedman (1982). See also Rochefort-Maranda (2011).

theory. Rather, partial interpretation as defined in the positivist sense above goes along with a certain commitment as to which purely theoretical concepts (such as electrons) refer. Now, as one normally does not want to give up the logical closure of statements of belief, the problem should be evaded by either switching from the syntactic view to the semantic view of theories, in particular by making use of the 'partial structure approach[9] or by taking a fictionalist stance on theoretical terms on which we can talk about protons, neutrons, electrons etc. as if they existed in the world without committing that they actually do (see Rochefort-Maranda (2011)). In the following, I will implicitly opt for the latter option.

None of the above is to say that the idea of concrete interpretations should or can only be cashed out by correspondence rules (on the syntactic view) or partial structures (on the semantic view):

- Uninterpreted theoretical terms may be linked to already interpreted ones (internal interpretations). Among others, (1) Stein (1994) proposes connecting a formal theory to the empirical by schematising the observer within the theory, taking it to be straightforward how the observer should then itself be linked to the world. (2) de Haro and de Regt (2018) provide examples of how already interpreted elements can be used for an interpretation of uninterpreted elements provided that they stand in a relevant relationship (such as symmetry).

- Correspondence rules may employ an (approximate) correspondence of a theoretical term to that of another theory for which the linkage to the observable is readily established (external-theory interpretation). Typically, external theories *to which the theory in question knowingly reduces* are used for this purpose. The so-called Schwarzschild mass in the Schwarzschild metric, for instance, can be to some degree interpreted by identifying its role as mass within the Newtonian limit theory.

2.1.1 The dynamical nature of the concrete interpretation

So far it may have sounded as if the concrete interpretation was easily fixed once and for all for a theory. However, the concrete interpretation of a theory — qua

[9] On the partial structure approach, the model is split up into a part which is known, and a part which is not known to refer. This allows for denoting the observable-part to refer and the un-observable to not refer. See Bueno (1997), section 3.

activity — is, in fact, an ongoing affair. It is, for instance, in this sense that we can read the following passage by Carnap (1966) on correspondence rules:

> Of course, physicists always face the danger that they may develop correspondence rules that will be incompatible with each other or with the theoretical laws. As long as such incompatibility does not occur, however, they are free to add new correspondence rules. The procedure is never-ending. There is always the possibility of adding new rules, thereby increasing the amount of interpretation specified for the theoretical terms; but no matter how much this is increased, the interpretation is never final. (p. 238)

Similarly, if you just loosely commit to cashing out the notion of concrete interpretation in terms of the contents of a 'good textbook' — the meme Curiel uses to display what he means by concrete interpretation — it should be clear that there is no definitive, final textbook to be expected on a topic like GR anyway.

In the following, I will make use of what I call empirical access to and empirical content of a theory which I define using the theoretical-observational distinction, and the notion of correspondence rules: The *empirical access* to a theory is given by the correspondence rules between theoretical and observational terms. It is increased when correspondence rules between a theoretical vocabulary and the observational vocabulary are added. The *empirical content* of a theory is given by the observational terms which are actually linked up to the theoretical terms by correspondence rules. It is increased when correspondence rules between a theoretical vocabulary and specific observational vocabulary which had neither directly nor indirectly (that is via other observational terms) been linked yet to theoretical terms become established. Increase in empirical content always implies an increase in empirical access but not vice versa.

The problem with the notion of empirical content is that it is highly dependent on what we take the observational vocabulary to be, which itself is, however, a problematic notion. As Carnap (1966) remarked,

> To a philosopher, "observable" has a very narrow meaning. It applies to such properties as "blue," "hard," "hot." These are properties directly perceived by the senses. To the physicist, the word has a much broader meaning. It includes any quantitative magnitude that can be measured in a relatively simple, direct way. A philosopher would

> not consider a temperature of, perhaps, 80 degrees centigrade, or a weight of $93\frac{1}{2}$ pounds, an observable because there is no direct sensory perception of such magnitudes. To a physicist, both are observables because they can be measured in an extremely simple way. The object to be weighed is placed on a balance scale. The temperature is measured with a thermometer. The physicist would not say that the mass of a molecule, let alone the mass of an electron, is something observable, because here the procedures of measurement are much more complicated and indirect. (pp. 225-226)

So, we will mainly focus on the notion of empirical access in the following. But even here one might be sceptical: When adding correspondence rules linking theoretical terms from T to observational terms from O via making recourse to external vocabulary from T' it is important to make sure that we are not just artificially inserting notions from T' in-between notions from T and O. What we have to require, thus, is that these additional paths from notions of T to notions of O via notions of T' feature notions of T' in a non-redundant fashion of one form or the other.

2.1.2 GR's concrete interpretation

A minimal concrete interpretation of GR includes the (potential) association of null geodesics with light rays, timelike curves with point particles of positive mass, and timelike geodesics with free point particles of positive mass (see, for instance, the interpretive principles (C1), (C2), and (P1) in Malament (2012), p. 120-121).

The, arguably, most minimal scheme for empirical access to the metric field in GR, then, builds on what is called the causal-inertial method: Based on a theorem by Weyl (1921)[10], Ehlers et al. (2012) provide a prescriptive scheme for determining the metric from the movement of light rays (linked to what's known as conformal structure) and freely falling particles (linked to what's known as projective structure) alone: an observer hereby tracks light rays and freely falling particles relative to local radar coordinates which use an arbitrary parameterisation of the observer's worldline as time parameterisation. It is important to stress that the scheme assumes that sufficiently well-defined radar coordi-

[10]The theorem basically states that "the projective and conformal state of a metric space determine the metric uniquely" (see Coleman and Korte (1980), Theorem 4.3).

nates can be set up in the first place.[11] Once the observer is thus schematised in the theory (see Stein (1994)), the metric structure can be interpreted internally through the procedure on how the observer measures out the metric structure using just the few elements of the theory already linked to the world (null geodesics, and time-like paths).

The causal-inertial method for accessing the metric field provides an internal interpretation of the metric field based on the minimal concrete interpretation of null geodesics and timelike curves. Similarly, one could adhere to a theorem by Fletcher (2013) for arguing that null geodesics (corresponding to light) as well as timelike curves (corresponding to massive particles) can be used to construct clocks *within the general relativistic model that read out the worldline interval.* Alternatively, one might, however, also interpret the metric structure through correspondence rules directly. The worldline interval, for instance, is linked by brute force to the reading of an ideal clock on the chronometric approach of Synge (1960).

It is also worth noting that, in regimes of weak gravitation, empirical content of GR can in a limited sense be fixed through Newtonian and post-Newtonian approximation of GR (external interpretation); as already mentioned before, the so-called Schwarzschild mass, for instance, acts effectively as a Newtonian point mass on observers who are far enough away from the inner region of the Schwarzschild spacetime.

3 Guidelines from GR's empirical interpretation

We can use the concrete interpretation of GR in at least two ways to get a glimpse at what kind of theory succeeds GR: (1) We can take the usually attributed empirical content seriously at a general level, that is wonder what other external theory-descriptions of the established empirical content there is, and whether these external theory-descriptions should not be merged with GR's formalism. This straightforwardly suggests a change of the formalism, and thus provides a direct guideline for theory development. (2) We can take the empirical content of the theory seriously at a more specific level: (a) We can consider whether — also under adherence to external theories — elements of GR's formalism should, after all, be empirically interpreted, that is linked to the

[11] The scheme was subsequently heavily improved; Coleman and Korte (1980), in particular, managed to free the scheme from charges of circularity. See Bell and Korté (2016) for a summary.

empirical by some correspondence rule. Call the posit behind this the *principle of maximal concrete interpretation* (PMCI): A theory's potential concrete interpretation should be exploited to a maximum degree, that is one should strive for the maximal (in principle) empirical content associable to a given theoretical formalism even if this may involve taking into account extratheoretical elements into its semantics. (b) We can consider whether — under adherence to external theories — the empirical content of GR becomes accessible in new ways. Call the posit behind this the *principle of maximal empirical access* (PMEA): One should strive for as many (in principle) modes of access possible to the empirical content even if this may involve taking into account extra-theoretical elements into its semantics. Positive results for (a) and (b) both *suggest* merging the external theories' formalism adhered to with GR's formalism. More precisely, when extending the standard theory's empirical interpretation through invoking an external theory, both the theory to be interpreted more and the theory invoked for this become to a certain extent entangled at the level of empirical interpretation: their empirical interpretations are not independent of each other anymore but at least one interpretation now also does recourse to the other. Such a strong entanglement at the semantic level then also suggests some sort of merging of the theories at the formal level. I now illustrate these rather abstract strategies (1) and (2). Thereby, I make two natural heuristics for motivating theory change into a particular direction explicit which are both based on quasi-empirical consideration.

3.1 General approach

GR-matter fields[12] $\phi_1, \ldots, \phi_n$ such as the electromagnetic field-strength tensor F_{ab} are linked to elements in the observable regime which — provided that curvature effects are negligible — are known to be locally more precisely described as quantum fields in flat spacetime than classical fields in curved spacetime. Thus,

- GR should be enhanced to take into account that the fields $\phi_1, \ldots, \phi_n$ are locally better described as quantum fields in flat spacetime than as classical fields (provided that local curvature effects are negligible). This

[12] In the standard formalism GR is formulated as a theory of fields on a 4-dimensional manifold with a symmetric rank-two tensor, the metric g of Lorentzian signature, and other (tensor) fields $\phi_1, \ldots \phi_m$ (such as the electromagnetic field-strength tensor F_{ab}) 'on top'. The fields $\phi_{i_1}, \ldots \phi_{i_n}$ which contribute to the energy-momentum tensor T are then called matter fields.

leads to QFT in curved spacetime, which treats quantum field theories in curved but static Lorentzian background geometries; and, as QFT in curved spacetime neglects back-reaction of the field to its now quantum matter content, to semi-classical gravity, which builds around the semi-classical field equations $G_{ab} = 8\pi\langle\hat{T}_{ab}\rangle$ where $\langle\hat{T}_{ab}\rangle$ is the expectation value of an operator-valued energy-momentum tensor (see Wald (1994)).

- The field equation links the field g (via the Einstein tensor G_{ab}) to the matter fields $\phi_1, \ldots, \phi_n$ (via the energy-momentum tensor T_{ab}), i.e. $G_{ab} = 8\pi T_{ab}$. The referents of the right-hand side of this equation are — as just stressed — known to be locally modelled as quantum theories in flat spacetime (provided that local curvature effects are negligible). This suggests then that G on the left-hand side, or rather g making up G, has referents that should be formulated as a quantum theory as well, and thus the quantisation of GR as such — at least if the semi-classical account turns out to insufficient.

3.2 Specific approaches

I present two examples — one from chronometry, the other from black hole thermodynamics — of how the concrete interpretation of GR can be extended under adherence to the (PMEC) or the (PMEA). Only the second of the two examples, however, provides a relevant insight as to where a successor theory to GR should be heading, which shows us that not every increase in empirical content or empirical access provides non-trivial hints.

3.2.1 Chronometric interpretation

Empirical access to the world-line interval in GR can be largely extended through the purely interpretative (albeit not necessarily empirically adequate) stipulation that external clocks — not modellable within the GR-framework itself — read out the worldline interval of the path they are travelling along (i.e., fulfil what has become known as the clock hypothesis[13]).[14]

[13]See Maudlin (2012), chapter 5, and Fletcher (2013).

[14]It is important to note, however, that *external* clocks are not required in *many* spacetimes for gaining chronometric access to the metric field: As a theorem by Fletcher (2013) demonstrates, light clocks internal to GR can be set up to measure the world-line interval of the metric up to arbitrary precision — provided that light can be said to move on null geodesics (see section 2.2.1).

How well this stipulation works, can be checked (for instance) using an operationalist criterion by Perlick (2008). So, whether a clock is a standard clock or not, is tested by tracking the movement of freely falling particles and light in (local) radar coordinates that are induced by the clock under consideration. More precisely, the clock in question is used to provide a parametrisation for the observer's world-line. Using this parametrisation and light signals, the observer can set up radar coordinates: Objects away from the word line are assigned a radial distance of $R = 1/2(t_2 - t_1)$ and a time $T = 1/2(t_2 + t_1)$ where t_1 is the parameter value on the world-line which corresponds to emission, and t_2 is the parameter value on the world line which corresponds to detection of the probing light signal. A clock reads out proper time at clock time t_0 (that is, is a standard/ideal clock) if and only if

$$\frac{\frac{d^2R}{dT^2}}{1-(\frac{dR}{dT})^2}\Big|_{t=t_0} = \pm\frac{\frac{d^2R'}{dT'^2}}{1-(\frac{dR'}{dT'})^2}\Big|_{t=t_0} (*)$$

where (R, T) and (R', T') denote the radar coordinates for two free particles respectively, sent off at $t = t_0$ into the same direction with differing speed ('+') or into opposite directions with arbitrary speeds (apart from that at least one needs to have a non-zero speed) ('-'). In other words, the extent to which a clock approximates a standard clock 'at a point' can be measured through the difference $\Delta = \left\|\left\|\frac{\frac{d^2R}{dT^2}}{1-(\frac{dR}{dT})^2}\Big|_{t=t_0}\right\| - \left\|\frac{\frac{d^2R'}{dT'^2}}{1-(\frac{dR'}{dT'})^2}\Big|_{t=t_0}\right\|\right\|$ with $\|..\|$ being a suitably chosen norm. For a freely falling standard clock, even $\frac{\frac{d^2R}{dT^2}}{1-(\frac{dR}{dT})^2}\Big|_{t=t_0} = 0$, i.e. $\frac{d^2R}{dT^2}\Big|_{t=t_0} = 0$. That is, the radar coordinates induced by a freely falling standard clock make the connection coefficients for the radial acceleration equation vanish at $t = t_0$.

Why should, however, a device fulfil $(*)$? That some external element (like an atomic clock) not at all describable in a general relativistic framework can be found to fulfil this condition, might be argued for from some selection process: we have simply managed to identify an 'apparatus' (such as an atom's oscillatory states) which suits our needs. The question of what ultimately makes this selection process possible in the first place remains. Now, the best explanation for why the atomic clock can be used to measure out the world-line interval is then simply that there is some sense in which quantum field theory (describing the atomic clock) and GR (describing the world-line interval) go along with each other. The hereby suggested unification is a bit sobering, however: we already know much more straightforwardly — namely from that the matter content adhered to in GR usually have a quantum description in approximately locally flat spacetime regions — that some kind of theory like QFT in curved spacetime/semi-classical gravity is needed (see section 3.1).

(As an alternative to the straightforward chronometric interpretation, one could — in tradition of the dynamical approach to relativity[15] stipulate the strong equivalence principle (SEP) — loosely speaking the idea that the matter field dynamics in GR is somewhat locally Minkowskian — as a means of establishing operational access to the metric via theory-external clocks:

> ... [the strong equivalence principle] allows us to carry over certain interpretational possibilities from SR. In particular, it allows us to transfer the interpretation of rods and clocks as waywisers of the metric tensor from the special case of the Minkowski metric to the case of a generically curved (but locally Minkowskian) metric, and it allows us to interpret the frames of reference in which the metric is locally Minkowskian as local inertial frames in the sense of 'inertial frame' we are wont to use from SR. The local validity of SR allows a 'trickling up' of interpretations from SR to GR. I said that this makes the role of the SEP seem interpretational, but we have to be careful not to see its role as 'merely' interpretational. The SEP explains why rods and clocks can serve as waywisers of the metric field. (Lehmkuhl (2011), p. 26)

The problem with this is that currently known versions of the SEP are not satisfactorily formulated to this purpose (see appendix).)

3.2.2 Thermodynamic interpretation

When seen within the narrow context of GR, the so-called black hole thermodynamic laws are statements on the relationship between geometry, and the energy and charges at asymptotic infinity of the spacetimes to which they apply; they are analogous to the standard thermodynamic laws (see table 1). However, once understood in the context of semi-classical GR (which clearly goes beyond GR proper)—in particular in light of Hawking's derivation of what is now called Hawking radiation—they are typically interpreted as proper laws of

[15]See Brown (2005), and, relatedly, Knox (2013).

	Standard thermodynamics	**Black hole thermodynamics**
Zeroth law:	T constant on a body in thermal equilibrium	κ constant on the horizon of a stationary black hole
First law:	$dE = TdS - pdV + \mu dN$	$dM = \frac{\kappa}{8\pi G} dA + \Omega_H dJ + \Phi_H dQ$
Second law:	$dS \geq 0$	$dA \geq 0$
Third law:	$T = 0$ cannot be reached	$\kappa = 0$ cannot be reached

Table 1: Overview of the analogies between standard thermodynamic laws and stationary black hole thermodynamics (largely taken from Kiefer (2004), p. 202). On the thermodynamic side, T is the temperature, S the entropy, p the pressure, V the volume, μ the chemical potential, and N the particle number of the standard thermodynamic system. On the black hole side of the analogy, κ is the surface gravity, A is the horizon area, Ω_H the angular velocity, J is the angular momentum, Φ is the electrostatic potential, and Q the electric charge of a black hole.

thermodynamics.[16]

But what keeps us from *interpreting* the surface gravity[17] starring in the black hole thermodynamic laws as thermodynamic temperature already at the level of GR proper? This association is (in principle) falsifiable, and well-motivated from the analogy (never mind that classical black holes do not allow for escape from its internal region, and are thus associated, if at all, with a temperature $T = 0$).[18] I consider now in what senses GR's black hole thermodynamic laws call for a tentative extension of GR's concrete interpretation.

More precisely: GR has a theoretical vocabulary T_{GR}, and thermodynamics that of $T_{\text{Thermodynamics}}$. The theory-overarching observational vocabulary is

[16] At least in the physics community. For critical accounts on the status of black hole thermodynamics qua thermodynamics, see Wüthrich (2019) and Dougherty and Callender (Forthcoming); for a defence of the orthodox view on black hole thermodynamics as more than an analogy, see Wallace (2017).

[17] Or a multiple thereof. After all, the first law only allows for an association of T with κ and S with A up to a positive constant k, i.e. $T = k\kappa$ and $S = \frac{A}{k}$.

[18] In semi-classical gravity this association is derived as what is known as Hawking radiation. For the original paper, see Hawking (1975). It was this finding which made physicists finally believe that black hole thermodynamics is more than an analogy. The point here is that this association would be worth trying *even if we had no hints from a theory beyond GR for its potential validity yet.* For a consequent albeit heavily heterodox plea for why even classical black holes should be assigned temperature (and in general, thermodynamic features), see Curiel (2014).

O. Correspondence rules $C(T_{\text{GR}}, O)$ connect elements from T_{GR} to elements in O. Notably, no correspondence rules on the standard conception links any theoretical term from GR to the "sensation of temperature", $o_{\text{heat sensation}} \in O$. Now, stipulate that the surface gravity $\kappa \in T_{\text{GR}}$ corresponds to the temperature $\in T_{\text{Thermodynamics}}$, that is that there is an inter-theoretical relation "$\kappa = c \cdot T$" $\in C'(T_{\text{GR}}, T_{\text{Thermodynamics}})$ where $C'(T_{\text{GR}}, T_{\text{Thermodynamics}})$ denote correspondence rules between the theoretical vocabulary of GR and that of thermodynamics, and c is a constant. Under this assumption, $\kappa \in T_{\text{GR}}$ could be associated to (1) the point coincidences observed on a thermometer $o_{\text{point coincidence}} \in O$, and arguably even to (2) the sensation of heat $o_{\text{heat sensation}} \in O$.

If case (1) is empirically established, *the empirical access linked to GR's empirical content is increased*, as now additional correspondence rules between κ and the observation of point coincidences exist (which run, in a non-redundant fashion, via the usage of a thermometer).

If case (2) was empirically established, even the empirical content linked to GR's formalism would be increased as correspondence rules between the surface gravity in GR and the sensation of heat are established for the first time. Note though that there seems to be an in principle limit to our human sensation of temperature — to still under-exaggerate — far off from the sensational sensitivity needed.

The problem, however, is that we expect from the QFT derivation of Hawking radiation that $T = \frac{\kappa}{8\pi}$ is too small for detection[19], not to say directly perceivable by a human observer.[20]

What is the thermodynamic interpretation good for then? That the thermodynamic interpretation could in principle increase empirical access to GR is a necessary requirement for that black hole thermodynamics can be rendered an instance of thermodynamics proper. Therefore, that we can at all *conceive* of the analogy as potentially increasing empirical access to GR, is a first good sign for that the analogy between black hole thermodynamics and thermodynamics proper is more than an analogy; in other words, our intuitive trust, if any, in black hole thermodynamics qua thermodynamics derives from the conceivability that thermometers, photon gases, ... can be brought into contact with a

[19]For astronomical black holes, the corresponding Hawking radiation is many orders smaller than the temperature of the cosmic microwave backgrounds. Small enough black holes for detection, on the other hand, would evaporate away too fast to be measurable.

[20]From the analogy at the level of GR, the putative black hole temperature can only be determined as equal to the surface gravity up to a constant pre-factor, which leaves the magnitude of the temperature undetermined.

black hole.[21] But this means that it is also *considerations of increased empirical accessibility* of GR via thermodynamics which foster the idea that black hole thermodynamics is thermodynamics proper.

4 Guidelines from GR's conceptual interpretation

Practice in quantum gravity research — the search for a successor theory to GR — goes far beyond motivating successors to GR from exploring GR's empirical content. Different quantum gravity approaches are rather suggested from stressing specific conceptual takes on GR, that is: Issues of categorial interpretation — Is the theory a geometrised spacetime theory? Is the theory a hydrodynamic theory? Is the theory a causal theory? — provide different perspectives on the theory; these conceptual categorisations of GR then suggest natural starting points for extrapolating GR, say from how these conceptual categorisations are already known to be linked to certain specific extrapolative strategies in other contexts.[22] For better illustration, let me provide some *examples* for links between conceptual interpretational stances on GR and extrapolative hypotheses:

- The categorisation of GR as a geometric spacetime theory directly derives from its standard differential geometric presentation (see, for instance, the canonical textbooks by Wald (2010) or Misner et al. (2017)). It has been explicitly voiced as such, among others, by Friedman (2014) and Maudlin (2012).[23] In giving the gravitational interaction a special character as intrinsically geometric, the geometric viewpoint naturally motivates a semi-

[21] See also Prunkl and Timpson (2017).

[22] The two examples invoked for demonstrating how enlarging the concrete interpretation drives theory change can also be thought of as special cases of categorial interpretations, namely categorial interpretations which have (in-principle) empirical testability: In the case of demanding that the surface gravity can also be measured via thermometers, GR is conceptualised as a (partly) thermodynamic theory, and, in the case of enriching the chronometric access to GR, GR is conceptualised as a (locally) special relativistic theory in a sufficiently strong sense (not just the dynamical equations are locally Lorentz-invariant but they are form-invariant).

[23] The alternative position is that of the field view, which loosely speaking, renders the metric field as just one field, among others. Of course, the metric field has special properties but so does every other field; the basic methodological posit behind the field view is then not to mistake (arguably contingent) matters of representation for decisive facts. The particle physics approach to GR ("spin-2") as, for instance, promoted by Weinberg and Dicke (1973) (see Salimkhani (2017) for a philosophical introduction), and hydrodynamic viewpoints on GR (see below) promote exactly such kind of take. Within the philosophy of physics, the field view is first and foremost promoted by adherents to the dynamical approach (see Brown (2005) as the locus classicus).

classical viewpoint according to which the geometric nature of spacetime is conserved and becomes the arena for quantum fields (semi-classical gravity).[24]

- GR can be seen as a locally thermodynamic theory: among other things[25], the field equations can be interpreted as a balance equation of a heat flux through the horizon of a (local) Rindler observer (see Jacobson (1995)).[26] Taking the thermodynamic viewpoint seriously — including the stipulation of a generalised second law of thermodynamics[27] — entails the holographic principle, which itself suggests a hypothesis on the number of microstates within a volume as being bounded by their surrounding area (see, for instance, Bousso (2002), section III).

- Connected to the previous viewpoint, GR can be seen as a hydrodynamic theory: under restriction to spacetimes with at least one Killing symmetry, the field equations take the form of the Navier-Stokes equations.[28] Rendering the field equations as hydrodynamic equations suggests considering fluctuation corrections to the field equations. This is exactly what is done in Hu's stochastic gravity (see Hu (1999)).

- Following a theorem by Malament (1977), spacetime structure in GR can be split up into (continuous) causal structure and local volume information. By making the causal viewpoint central, the volume information can be demoted to a secondary feature. (Provided that we assume a finite number of causal events, the volume information can then be obtained through counting the number of causal events in a region.) This naturally leads to causal set theory.

- GR is universal coupling. This can be seen either as a necessary feature of its putative geometric nature (see first point), or as a sign for that it is the result of coarse-graining (its geometric nature is then rather a representational coincidence).[29] The latter view suggests taking analogies

[24] For a discussion of arguments against semi-classical accounts as fundamental theories, see Callender and Huggett (2001) and Wüthrich (2005). For a modern thought experiment for testing the scope of a semi-classical gravity paradigm, see Bose et al. (2017).

[25] See Padmanabhan (2016), section 1 for a comparison of various account of gravity as a thermodynamic theory.

[26] Strictly speaking, this requires, however, adherence to the Unruh effect, that is an effect from QFT in flat spacetime.

[27] The second law holds for matter and black hole entropy in total.

[28] See, for instance, Rodrigues Jr. and de Oliveira (2016), chapter 15.

[29] See Feynman et al. (2003), section 1.5.

between GR and non-geometrical theories more seriously again.

We can note that, in practice, many principles and associated viewpoints just seem to make *themselves* remarkable, say by analogy to other theories. However, one way to systematically work out decisive principles and thus viewpoints of GR is a contrastive approach where GR is compared to neighbouring spacetime theories in order to reveal features making GR special, and thus worth specific attention — see Lehmkuhl et al. (2017) for a project along these lines.

Furthermore, we can note that each conceptual take on GR above can usually be backed up from several strands of reasoning. It is an urgent question then for the philosophy of quantum gravity qua philosophy of discovery to what extent robustness arguments which work at an entirely conceptual and thus non-empirical level can support the pursuit-worthiness or even plausibility of a specific conceptual viewpoint on GR.

Concerning the motivation of hypotheses from a specific viewpoint, we can make the following two observations: (1) Each of these viewpoints on GR suggests characteristic extrapolative hypotheses. In many cases, these hypotheses are suggested by analogy: a particular viewpoint is known to be linked to specific extrapolative strategies in the context of other theories, which can be exploited in the context of GR then. In some other cases, taking a viewpoint seriously more or less requires making new technical commitments or discarding old ones. Causal set theory is a good example for this as it arises from taking the causal viewpoint seriously which means discretising the space of events, and thus giving up a notion of Lorentz-symmetry and Lorentzian manifold at high energies.[30]
(2) By suggesting characteristic extrapolative hypotheses, categorial interpretations obtain a decisive role in motivating prima facie independent approaches. What we call the principles of a specific approach to quantum gravity, is thus often already rooted in a specific way of looking at GR, i.e. a specific categorial interpretational stance.

5 Conclusion

This essay showed how both empirical and conceptual interpretations of GR give straightforward suggestions for a successor theory to GR. I identified the following heuristic rules for driving theory change from interpreting GR:

[30] Which does not mean that CST is inconsistent with Lorentz symmetry at lower energies. See Dowker et al. (2004).

- At the level of the empirical interpretation:

  (1) Take the empirical content seriously at a general level: Look for entities adhered to in the theory but knowingly described more accurately in certain domains by other theories. This suggests potential directions for unification.

  (2) Take the empirical content seriously at a specific level: Look for extratheoretical measurement methods of otherwise empirically uninterpreted parts of the theory. This suggests potential directions for unification of GR.

- At the level of the conceptual interpretation:

  (1) Explore possible conceptual readings of the theory. Directions for conceptual readings are at first suggested by striking conceptual analogies to other theories, and prominent conceptual features or results as such.

  (2) Try extrapolative schemes generally associated to a certain conceptual categorisation of the theory.

  (3) If a conceptual viewpoint is novel, that is not known from other theories (as the viewpoint of GR as a causal theory), explore options for making this viewpoint centre-stage nevertheless. This may easily involve violating principles of the current theory (such as that of a Lorentzian manifold in the case of causal set theory). In fact, possible conflicts again drive progress here, as they provide a concrete problem to tackle.

Now, the heuristic of using empirical and concrete interpretation as a guide towards successor theories can, of course, again be used on the thereby suggested successor theories such as QFT in curved spacetime, semi-classical gravity, and quantum GR. The full merit of the promoted methodology thus derives from its iterative applicability.

A final remark is in order on the relationship between theory interpretation and internal problems as theory drivers: The essay investigated how theory interpretation suggest directions for theory change. At the same time, it is a common theme — in particular in quantum gravity research where the problem is first of all theoretical and not empirical[31] — that internal problems suggest the alteration of a theory into a certain direction. Now, it is worth stressing that of course not the problem as such but rather the conceptual interpretation of the theory featuring the problem suggests the direction for theory change. The problem as such can at most only indicate the need for theory change.

References

John L. Bell and Herbert Korté. Hermann Weyl. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, winter 2016 edition, 2016.

Sougato Bose, Anupam Mazumdar, Gavin W. Morley, Hendrik Ulbricht, Marko Toroš, Mauro Paternostro, Andrew A. Geraci, Peter F. Barker, M.S. Kim, and Gerard Milburn. Spin entanglement witness for quantum gravity. *Phys. Rev. Lett.*, 119(24):240401, 2017.

Raphael Bousso. The holographic principle. *Reviews of Modern Physics*, 74(3): 825, 2002.

Harvey R. Brown. *Physical relativity: Space-time structure from a dynamical perspective*. Oxford University Press, 2005.

Otávio Bueno. Empirical adequacy: A partial structures approach. *Studies in History and Philosophy of Science Part A*, 28(4):585–610, 1997.

Craig Callender and Nick Huggett. *Physics meets philosophy at the Planck scale: Contemporary theories in quantum gravity*. Cambridge University Press, 2001.

Rudolf Carnap. *Philosophical foundations of physics*, volume 966. Basic Books New York, 1966.

Robert Alan Coleman and Herbert Korte. Jet bundles and path structures. *Journal of Mathematical Physics*, 21(6):1340–1351, 1980.

Erik Curiel. General relativity needs no interpretation. *Philosophy of Science*, 76(1):44–72, 2009.

[31] Examples include: (1) GR spacetimes are partly singular. These spacetimes are not sufficiently predictive. (2) QFT in curved spacetime lacks back-reaction (which we know is desirable from GR). (3) Semi-classical gravity is possibly incoherent and thus at least predictively limited. (4) Perturbative quantisation of gravity leads to an only effectively renormalisable, that is predictively limited theory.

Erik Curiel. Classical black holes are hot. *arXiv preprint arXiv:1408.3691*, 2014.

Sebastian de Haro and Henk W. de Regt. Interpreting theories without a spacetime. *European Journal for Philosophy of Science*, pages 1–40, 2018.

John Dougherty and Craig Callender. Black hole thermodynamics: More than an analogy? In Barry Loewer, editor, *Philosophy of Cosmology*. Cambridge University Press, Forthcoming. http://philsci-archive.pitt.edu/13195/.

Fay Dowker, Joe Henson, and Rafael D. Sorkin. Quantum gravity phenomenology, lorentz invariance and discreteness. *Modern Physics Letters A*, 19(24): 1829–1840, 2004.

Jürgen Ehlers, Felix A. E. Pirani, and Alfred Schild. Republication of: The geometry of free fall and light propagation. *General Relativity and Gravitation*, 44(6):1587–1609, 2012.

Richard P. Feynman, Fernando B. Morinigo, and William G. Wagner. Feynman lectures on gravitation, 2003.

Samuel C. Fletcher. Light clocks and the clock hypothesis. *Foundations of Physics*, 43(11):1369–1383, Nov 2013. ISSN 1572-9516. URL https://doi.org/10.1007/s10701-013-9751-3.

Michael Friedman. The scientific image by Bas C. van Fraassen. *The Journal of Philosophy*, 79(5):274–283, 1982.

Michael Friedman. *Foundations of space-time theories: Relativistic physics and philosophy of science*, volume 113. Princeton University Press, 2014.

Hans Halvorson. The semantic view, if plausible, is syntactic. *Philosophy of Science*, 80(3):475–478, 2013.

Stephen W. Hawking. Particle creation by black holes. *Communications in mathematical physics*, 43(3):199–220, 1975.

B. L. Hu. Stochastic gravity. *International Journal of Theoretical Physics*, 38 (11):2987–3037, 1999.

Ted Jacobson. Thermodynamics of spacetime: the Einstein equation of state. *Phys. Rev. Lett.*, 75(7):1260, 1995.

Claus Kiefer. Quantum gravity. *Int. Ser. Monogr. Phys.*, 124:1–308, 2004.

Eleanor Knox. Effective spacetime geometry. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, 44 (3):346–356, 2013.

Dennis Lehmkuhl. Mass–energy–momentum: Only there because of spacetime? *The British Journal for the Philosophy of Science*, 62(3):453–488, 2011.

Dennis Lehmkuhl. Literal versus careful interpretations of scientific theories: The vacuum approach to the problem of motion in general relativity. *Philosophy of Science*, 84(5):1202–1214, 2017.

Dennis Lehmkuhl, Gregor Schiemann, and Erhard Scholz. *Towards a theory of spacetime theories*, volume 13. Springer, 2017.

Niels Linnemann. On interpreting the empirical content of general relativity. *Unpublished draft*, 2019.

Sebastian Lutz. What's right with a syntactic approach to theories and models? *Erkenntnis*, 79(8):1475–1492, 2014.

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

David B. Malament. The class of continuous timelike curves determines the topology of spacetime. *Journal of mathematical physics*, 18(7):1399–1404, 1977.

David B. Malament. *Topics in the foundations of general relativity and Newtonian gravitation theory*. The University of Chicago Press, 2012.

Tim Maudlin. *Philosophy of physics: Space and time*, volume 5. Princeton University Press, 2012.

Charles W. Misner, Kip S. Thorne, and John Archibald Wheeler. *Gravitation*. Princeton University Press, 2017.

Thanu Padmanabhan. Exploring the nature of gravity. *arXiv preprint arXiv:1602.01474*, 2016.

Volker Perlick. On the radar method in general-relativistic spacetimes. In *Lasers, Clocks and Drag-Free Control*, pages 131–152. Springer, 2008.

Carina Prunkl and Christopher Timpson. Black Hole Entropy is Entropy and not (necessarily) Information. *Unpublished draft*, 2017.

Guillaume Rochefort-Maranda. Constructive empiricism and the closure problem. *Erkenntnis*, 75(1):61–65, 2011.

Waldyr A. Rodrigues Jr. and Edmundo Capelas de Oliveira. *The Many Faces of Maxwell, Dirac and Einstein Equations: A Clifford Bundle Approach*, volume 922. Springer, 2016.

Kian Salimkhani. Quantum Gravity: A Dogma of Unification? *Philosophy of Science – Between Natural Science, the Social Sciences, and the Humanities, European Studies in Philosophy of Science, Springer*, 2017.

Howard Stein. Some reflections on the structure of our knowledge in physics. In *Logic, Metholodogy and Philosophy of Science, Proceedings of the Ninth International Congress of Logic, Methodology and Philosophy of Science*, pages 633–55. Citeseer, 1994.

Frederick Suppe. Understanding scientific theories: An assessment of developments, 1969-1998. *Philosophy of Science*, 67:S102–S115, 2000.

John Lighton Synge. *Relativity: the general theory.* North-Holland Publishing Company Amsterdam, 1960.

Bas C. Van Fraassen. *The scientific image.* Oxford University Press, 1980.

Robert M. Wald. *Quantum field theory in curved spacetime and black hole thermodynamics.* The University of Chicago Press, 1994.

Robert M. Wald. *General relativity.* The University of Chicago Press, 2010.

David Wallace. The case for black hole thermodynamics, Part i: phenomenological thermodynamics. *arXiv preprint arXiv:710.02724*, 2017.

James Owen Weatherall. Against dogma: On superluminal propagation in classical electromagnetism. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, 48: 109 – 123, 2014. ISSN 1355-2198. doi: https://doi.org/10.1016/j.shpsb.

2014.08.005. URL http://www.sciencedirect.com/science/article/pii/S1355219814000896. Relativistic Causality.

Steven Weinberg and R. H. Dicke. *Gravitation and cosmology: principles and applications of the general theory of relativity.* Wiley New York, 1973.

Hermann Weyl. Zur Infinitesimalgeometrie: Einordnung der projektiven und der konformen Auffasung. *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse*, 1921:99–112, 1921.

Rasmus G. Winther. The structure of scientific theories. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy.* Metaphysics Research Lab, Stanford University, winter 2016 edition, 2016.

Christian Wüthrich. To quantize or not to quantize: fact and folklore in quantum gravity. *Philosophy of Science*, 72(5):777–788, 2005.

Christian Wüthrich. Are black holes about information?, 2019. In Richard Dawid, Radin Dardashti, and Karim Thébault, editors, *Epistemology of Fundamental Physics*, Cambridge University Press, 2019.