Causation and the conservation of energy in general relativity

Forthcoming in *The British Journal for the Philosophy of Science*

Sebastián Murgueitio Ramírez, James Read, and Andrés Páez

Abstract

Consensus in the contemporary philosophical literature has it that conserved quantity theories of causation such as that of Dowe [2000]—according to which causation is to be analysed in terms of the exchange of conserved quantities (e.g., energy)—face damning problems when confronted with contemporary physics, where the notion of conservation becomes delicate. In particular, in general relativity it is often claimed that there simply *are* no conservation laws for (say) total-stress energy. If this claim is correct, it is difficult to see how conserved quantity theories of causation could survive. In this article, we resist the above consensus and defend conserved quantity theories from this conclusion, at least when focusing on the apparent problems posed by general relativity. We argue that this approach to causation can continue to be defended in general relativity, once one appreciates (a) the availability of approximate symmetries in generic general relativistic spacetimes, and (b) the role of modelling and idealisation in that theory. Given these points, conserved quantity theories of causation must stand or fall on other grounds.

1 Introduction

Physical connection theories of causation—which analyze causal interactions in terms of physical processes ‘connecting’ the cause and the effect—have been accused of being unable to handle causation in physics, especially in the context of general relativity (GR). In particular, Dowe’s [2000] conserved quantity theory of causation (CQTC), which is arguably the most developed physical connection theory, has been said to fail to make sense of causation in GR because CQTC requires the use of conserved quantities, and in GR it is famously difficult to identify suitable conservation laws (e.g., see Rueger [1998], Curiel [2000] and, more recently, Vassallo [2019]). And given that CQTC is supposed to be heavily motivated by physics (Dowe [2000], pp. 1–13), the alleged incompatibility between CQTC and GR is a rather serious problem for CQTC. To this day, there is no adequate response to the challenge posed by this alleged incompatibility. As Vassallo ([2019], p. 5) put it recently, ‘it is apparent that defending the conserved quantity approach in this context [GR] is a tremendously hard task.’

In this paper, by building on recent developments regarding conservation laws in GR and by paying close attention to the ways in which physicists use and interpret models in physical theories, we develop a response to this challenge. We show, in particular, that there are some sophisticated—and thus far unrecognized—means by which physical connection theories such as CQTC can make sense of causal processes and causal interactions in GR. In order to do so, we first present a brief recapitulation of Dowe’s theory (Section 2) followed by a very detailed reconstruction of the challenge (Section 3). Then, in Section 4, we put forward two different ways by which the notion of an approximate symmetry can be used in order to face the challenge. In Section 5, we present a yet different way of responding to the challenge, this time by defending the legitimacy of certain local conservation laws in GR against recent charges in the philosophical literature. We close by pointing out certain remaining difficulties a defender of CQTC might still face in the context of GR, and by drawing some general lessons regarding causal explanations in GR which are relevant to any philosophical approach that seeks to make sense of causal statements in concrete applications of GR.

2 Preliminaries: Dowe’s theory

Before we begin examining the challenge that GR poses to Dowe’s theory, it is convenient to present some of the basic claims of the conserved quantity theory of causation (CQTC). The two basic notions of CQTC are *causal processes* and *causal interactions*. Dowe ([2000], p. 90) defines them as follows:

CQ1: A *causal process* is a worldline of an object that possesses a conserved quantity.

CQ2: A *causal interaction* is an intersection of worldlines that involves exchange of a conserved quantity.

Dowe offers the following definitions of the terms 'intersection,' 'exchange,' 'conserved quantity,' and 'possesses' mentioned in CQ1 and CQ2. We label them for future reference:

D1: An *intersection* is simply the overlapping in spacetime of two or more processes. The intersection occurs at the location consisting of all the spacetime points that are common to both (or all) processes (pp. 91–92).

D2: An *exchange* occurs when at least one incoming and at least one outgoing process undergoes a change in the value of the conserved quantity, where 'outgoing' and 'incoming' are delineated on the spacetime diagram by the forward and backward light cones, but are essentially interchangeable. The exchange is governed by the conservation law, which guarantees that it is a genuine causal interaction (p. 92).

D3: A *conserved quantity* is any quantity that is governed by a conservation law, and current scientific theory is our best guide as to what these are. For example, we have good reason to believe that mass-energy, linear momentum, and charge are conserved quantities (p. 91).

D4: 'Possesses' is to be understood in the sense of 'instantiates.' An object possessing a conserved quantity is an instance of a particular instantiation of a property. We suppose that an object possesses energy if science attributes that quantity to that body. It does not matter whether that process transmits the quantity or not, nor whether the object keeps a constant amount of the quantity. It must simply be that the quantity may be truly predicated of the object (p. 92).

A simple example will show how these definitions can be used. Suppose asteroid A hits asteroid B, and the impact sends the latter in direction $\hat{\vec{r}}$ (in a given coordinate system). Also, suppose that these two asteroids are so far away from everything else that they constitute an isolated system. We know from Newton's laws that the total linear momentum is a conserved quantity if the total system (the two asteroids) is isolated. Therefore D3 is satisfied (in reality, what matters is not that the system in question is such that it has a quantity that is indeed conserved, but only that it has a quantity such that it would be conserved if the system were isolated). It can also be established that both asteroids *possess* conserved quantities in the sense of D4: physics can truly predicate linear momentum of both asteroids. Therefore, the worldlines of both asteroids are causal processes according to CQ1. Now, both worldlines *intersect* in the sense of D1: there is a time t at which the spatial regions occupied by asteroids A and B overlap. Furthermore, during the intersection an exchange of linear momentum occurs, in accordance with D2. During the exchange the linear momentum of both asteroids changes. Thus, according to CQ2, there is a causal intersection between the asteroids.

3 Clarifying the challenge

In a nutshell, the main challenge that GR poses for CQTC is that it seems that conservation laws cannot be formulated except for very special kinds of spacetimes—so special that our spacetime is excluded. Hence, it follows from CQTC that there cannot be causal processes (and therefore causal interactions) in our spacetime (for there would not be conserved quantities). But surely there are causal processes and causal interactions in our spacetime. Therefore, the challenge goes, CQTC must be false.

Now, in order to offer ways by which CQTC can face this kind of challenge, it is important to first present a more detailed formulation of it. And to do so we will take as a starting point the formulation given by Curiel ([2000]) (although we will present it in a slightly different manner). First, we note that we can divide conserved quantities into two kinds: (i) those that admit of both global (or integral) and local (or differential) conservation laws, and (ii) those that only admit of local conservation laws. Next, we offer some reasons why quantities of kind (ii) are pathological. We then show that one cannot define conservation laws for quantities of kind (i) (i.e., we cannot define integral conservation laws for these quantities) except under extremely special circumstances (circumstances that, crucially, do not apply to our spacetime). Hence, in our universe and in the vast majority of spacetimes described by GR, we cannot define adequate (non-pathological) conserved quantities. Finally, it follows from CQTC that there cannot be causal processes (and, consequently, causal interactions) in our spacetime because there simply are no (non-pathological) conserved quantities. In the following passage Curiel makes essentially this same point (after assuming that the only acceptable conservation laws are the integral ones):

> The spacetimes in which Killing fields occur, however, are highly special and unphysical. [...] The smallest speck of dust [...] in only one spot in the entire spacetime precludes the existence of a Killing field. [... I]t goes without saying that the actual spacetime we inhabit possesses no Killing fields. (Curiel [2000], pp. 47-48)

Other authors, such as Rueger ([1998]) and Lam ([2010]), have focused on the same kind of point: the fact that one can only have integral conservation laws in the presence of Killing fields means that, for many cases of interest (including our own universe), CQTC entails that there are no causal processes. At this point, it will be convenient to present this argument in a slightly more formal manner. As we construe it, the argument can be put as follows:

1. There are two kinds of conserved quantities: quantities that admit of both global (integral) and local (differential) conservation laws ('kind 1 quantities'), and quantities that admit only of local (differential) conservation laws ('kind 2 quantities').

2. If spacetime S lacks Killing vector fields, then S admits only kind 2 quantities.

3. Kind 2 quantities are pathological.

4. Our spacetime lacks Killing vector fields.

5. Hence, there are no non-pathological conserved quantities in our spacetime. (2–4)

6. CQTC requires non-pathological conserved quantities.

7. Hence, if CQTC is true, then there cannot be causal processes and causal interactions in our spacetime. (5, 6)

8. There are causal processes and interactions in our spacetime (e.g., the detection of gravitational waves by LIGO and Virgo).[1]

C. Hence, CQTC is false. (7, 8)

The argument (henceforth 'CHALLENGE') is valid, so if one wants to defend CQTC, one must find a premise to reject. The plan for the rest of the paper consists of exploring two main ways of resisting this argument. First, in Section 4, we will explore ways of resisting premise 4. Then, in Section 5, we will consider ways of undermining premise 3.

Before we move on, we want to emphasize that CHALLENGE aims to show that CQTC is false on the basis that it cannot capture a *single* causal process or a *single* causal interaction in our spacetime (premise 8 requires only that there be one causal process or interaction). Recall from the quote above that Curiel stresses that 'the actual spacetime we inhabit possesses no Killing fields' as a way of explaining why CQTC cannot capture (any) causal processes and causal interactions in our spacetime. Hence, CHALLENGE is rather strong, and it can be undermined by showing that CQTC can capture at least *some* causal processes or causal interactions in our spacetime, which is precisely the goal of the next sections. Similarly, as we will discuss soon, we do not commit to the stronger claim that CQTC captures every single *putative* causal process or interaction in spacetime, and there are interesting questions precisely on what the scope of applicability for CQTC is in the context of GR.

4 Approximations to the rescue

In replying to Rueger ([1998]), Dowe writes that 'As I understand it, our spacetime does exhibit the right symmetry; global conservation laws do hold in our universe as far as we know. I take it, then, that the conserved quantity theory is not refuted' ([2000], p. 97). As we see it, Dowe can be interpreted as rejecting premise 4 (i.e., as rejecting that our spacetime lacks Killing

[1] Here, 'causal processes' just means processes or objects capable of engaging in causal interactions. It cannot mean 'objects possessing conserved quantities', for that would assume CQTC's framework, which the argument seeks to reject.

vector fields) in the previous argument by asserting that our universe does indeed have Killing vector fields (notice that he says that it has the 'right symmetry').[2] Unfortunately, Dowe does not elaborate on this point, and what he says is insufficient. However, we think that recent developments in GR provide a natural way of elaborating on Dowe's remark that makes his attack on premise 4 much more promising, as we will next explain.

Consider the following proposal, consisting of three main claims. First, whether or not a given spacetime admits certain symmetries is a vague matter in the following sense: it can admit them exactly when spacetime has these symmetries exactly, or it can admit them approximately when spacetime has these symmetries approximately. Hence, even if our spacetime lacks exact Killing fields, that still leaves open a way of resisting premise 4 of the argument by suggesting that our universe has Killing fields (or other symmetries) in an approximate manner.[3] Second, suppose that this approximate admission of symmetries is enough for the purposes of deriving integral conservation laws (notice that this can be regarded as a way of rejecting premise 2 in CHALLENGE). And third, assume that the conserved quantities associated with those symmetries are of the right kind for the identification of causal processes and causal interactions along the lines of CQTC. Notice that either the first claim or the second claim is sufficient for resisting CHALLENGE (for they both offer ways in which to reject a premise). This is worth emphasizing because, as far as we can tell, ours is the first paper using approximate symmetries in a way to respond to CHALLENGE.[4] However, the third claim of the current proposal is important if we want to show that 'adding' approximate symmetries to spacetime does not bring about new problems for CQTC! In what follows, we will elaborate on this proposal, and show how it is supported in various forms by recent developments in GR.

Let's start by considering the question of what it means to say that a spacetime admits symmetries approximately. Intuitively, we can think of this notion of approximate symmetries as the idea that our spacetime admits fields that, in some domains and to certain degrees of precision, behave very similarly to Killing vector fields. Now, this intuitive notion has been developed in different ways in the physics literature; here we want to highlight two such ways. One way of making sense of approximate symmetries comes via the concept of 'almost Killing fields' (AKF from now on): fields whose flow preserves the metric to certain orders (exact Killing fields preserve the metric to all orders). This idea is not new; it can be traced at least to

[2]Consider a diffeomorphism d such that $d^{*}g_{ab} = g_{ab}$, where star denotes the push-forward map. In this case, said diffeomorphism is an *isometry* of the Lorentzian metric field g_{ab}. For every such isometry there is a Killing vector field and associated conserved quantity. In what follows, we assume that the reader is familiar with this standard story, but see (Read [2020]) for recent discussion in the philosophical literature, and (Wald [1984]) for a very elegant textbook presentation.

[3]Here, one might ask: what of GR spacetimes with no approximate symmetries? As a matter of fact, the constructions regarding approximate symmetries due to Harte ([2008]) which we will invoke in what follows hold in *all* GR spacetimes—so there should (at least potentially—more on this below) be a sense in which, on CQTC, there is causation in all such spacetimes.

[4]It is worth noting that Pitts ([2022a], p. 1964) says that 'given that approximate symmetries are a mainstream problem with a strong intuitive basis, it seems fairly plausible that an adequate solution [to CHALLENGE] exists and might well be found.' Hence, it seems that Pitts, like us, recognizes approximate symmetries as a promising way of defending CQTC in the context of GR—but, unlike us, he does not elaborate on the proposal.

the 1960s (Matzner [1968]), although it has been developed in various forms since then (e.g., Harte [2008]). Let's call this approach/program 'Almost Killing,' as it makes use of the concept of AKFs. The other way of making sense of approximate symmetries that we want to highlight appeals to the concept of asymptotically flat spacetimes: spacetimes that, as the name indicates, approach Minkowski's asymptotically. In particular, if there are subsystems of our universe that can be well approximated as being embedded in asymptotically flat spacetimes, then we could say that those subsystems exhibit (asymptotic) symmetries approximately. For brevity, let's call this proposal 'Asymptotic Symmetry.' We now consider these proposals in further detail.[5]

4.1 Almost Killing

As the name indicates, the central concept behind Almost Killing is that there are fields that, in some sense, behave very much like Killing fields. However, the sense in which they behave similarly to Killing fields varies from proposal to proposal. Here we draw particular attention to the work of Harte ([2008]), in which 'Jacobi fields' (understood as generalized affine collineations—the latter being the generators of infinitesimal affine transformations: see (Harte [2008], p. 3)) are used in order to construct a more general notion of a Killing vector field.[6] Crucially, these Jacobi fields correspond to exact symmetries along the worldline of an observer (i.e., along a timelike worldine), and they give rise to a precise notion of approximate symmetry near the wordline (Harte [2008], §3). And even though these approximate symmetries are local (in the sense that they are defined 'around' a worldline), they can be extended in a non-perturbative way to finite regions around the worldline from which they were constructed ([2008], p. 13).[7]

Importantly for us, these generalized affine collineations can also be used to define conserved quantities and 'almost conserved quantities' (quantities that vary very slowly: see ([2008], pp. 14–15)). Indeed, these Jacobi fields can be used to recover the standard conserved quan-

[5] A word on the scope of the arguments that follow. The notions of approximate/asymptotic symmetries which we consider apply not just to the actual world, but to all worlds (or parts of worlds) represented well by the models of GR. Thus, the constructions which we invoke afford the means to adjudicate on whether there is genuine causation in a given world (or part of a world). Of course, though, what are going to be of particular interest are the models of GR which represent (parts of) the *actual* world—for, ultimately, we are interested in whether there can be causation in the actual world.

[6] Philosophers have also engaged with the notion of approximate symmetries and approximate Killing fields in spacetime theories—see (Fletcher [2020]) and (Fletcher and Weatherall [2022]). Though interesting in itself, in this article we focus on the work of Harte ([2008]) in order to make our central points, mainly due to the fact that this work is at present more established in the physics literature. For some comparison of these works, see (Linnemann *et al.* [2023]).

[7] In this sense, they are 'quasi-local'. Although we choose to focus on Harte's construction in this article, it is worth stressing that there is a cornucopia of other quasi-local approaches to gravitational stress-energy and energy conservation in GR, of which the proponent of CQTC may be able to avail themselves (see (Szabados [2009]) for the *locus classicus*). We do not discuss these other approaches further in this article as Harte's construction is sufficient to make our points. Those unconvinced by Harte in particular, though, are invited to study these other approaches (though it is worth noting that many of such approaches also face conceptual issues: e.g., non-uniqueness, various ambiguities, etc.).

tities associated with exact Killing fields, but they go beyond that insofar as they also provide additional conserved quantities not derivable from exact Killing fields, and allow us to express other quantities that are not conserved but are close to being conserved as one approaches the worldline of an observer (along the wordline, these quantities are exactly conserved). In addition, Harte ([2008]) shows how to use these Jacobi fields to construct conservation laws for extended matter distributions (as opposed to point particles), and, in particular, he shows how to use these fields in order to construct an integral conservation law for a generalized version of the Komar momentum ([2008], p. 16). This last point is crucial, for it highlights that even in the absence of exact Killing fields, one can still appeal to AKF in order to construct integral conservation laws. Note that these constructions from Harte ([2008]) apply, at least in principle, to all GR spacetimes and they even work for regions of strong curvature (and hence far from the conditions where Killing fields normally appear), such as those encountered in the vicinity of a black hole.[8]

In short, there are four features of the AKF discussed by Harte ([2008]) that we wish to highlight here:

1. These symmetries always exist along a timelike worldline, even when the spacetime under consideration lacks isometries, and even in conditions of strong curvature.

2. These symmetries are exact along a worldline and they are approximate (in a precise sense) as one moves away from it.

3. Even though they are defined locally (at a wordline), these symmetries can be extended to finite regions in a non-perturbative way.

4. These symmetries can be used to compute integral laws for conserved quantities and almost conserved quantities (quantities that vary very slowly) for both particles and extended bodies.

Armed with these generalized symmetries, it is natural to suggest that they suffice for the fruitful application of CQTC even in the absence of exact Killing fields. To do so, recall condition D3 in CQTC: a conserved quantity is any quantity that is governed by a conservation law, and current scientific theory is our best guide to what these are. The suggestion, then, is to say that the conservation laws and the 'almost conservation laws' arising from these generalized symmetries should count within the scope of D3, and so can be used for defining causal processes and causal interactions. And given that we can associate these general symmetries to any wordline even in spacetimes that lack exact Killing fields (and even in cases of strong

[8] One has to be careful, however, when using some of the quantities defined by Harte in the case of black holes. In particular, for the framework to work, the differential equations must be well-defined, and so one would have to avoid, for example, $r = 0$ in the Schwarzschild solution (but if one moves to a nearby region, even if it has strong curvature, the approach will work). We are grateful to an anonymous referee for prompting us to say something on strong gravity regimes, and to Abraham Harte for discussion on the point.

curvature, as explained above), this opens the door for the possibility of CQTC recognizing causal processes in a much wider range of situations than previously suspected!

At this point, let us consider what we think might be one of the main criticisms of this kind of proposal, namely, that the extent to which it is compatible with CQTC is unclear. One might think, for instance, that CQTC has no room for almost-conserved quantities of the kind entailed by this proposal, perhaps because accepting these kinds of quantities might lead CQTC to the recognition of 'almost-causal processes' and 'almost-causal interactions,' notions that might not make much sense; either two objects undergo a causal interaction or they don't, one might think. However, we find this sort of worry to be a bit premature. The proposal in question is not to make the notion of causal interaction a vague one, but rather to extend the scope of CQTC by making the notion of 'conserved quantity' somewhat more flexible. So, under this proposal, if two processes instantiate almost-conserved quantities, and if such processes exchange certain amount of those almost-conserved quantities, then one would say that such processes had causally interacted (not that they 'almost causally interacted' or something along those lines). In other words, the proposal is to say that an interaction is causal if it involves the exchange of either conserved quantities or almost-conserved quantities, and it is not causal otherwise.

Here is another related and important point. Suppose that a quantity is (merely) almost-conserved, but the extent to which it is not conserved is irrelevant to the particular experimental/practical/measurement context under consideration (perhaps, for example, one's measurement devices are insufficiently sensitive to detect said non-conservation). In that case, it is reasonable to maintain—following the methodology of Wallace explicated famously in the context of classical worlds in the Everett interpretation of quantum mechanics (Wallace [2012])—that 'for all practical purposes' there is indeed conservation. Moreover, insofar as that (quasi-)conservation is robust and useful in certain contexts (cf. Wallace's discussion of what he calls 'Dennett's criterion' (Wallace [2012], p. 50)), one can maintain that there *just is* conservation in such cases: for (to return to the case of gravitational energy and conservation) in such cases there is something which plays the functional role of a conserved quantity such as gravitational energy (cf. (Read [2020])). Of course, this approach to the identification of emergent ontology and laws is not uncontroversial (for a well-known detractor, consider Maudlin [2019]), but what we wish to stress here is that those who embrace Wallace's use of 'for all practical purposes' will likely be particularly inclined to view 'almost conservation' as being unproblematic.

Here it might also be useful to recall that one of the main reasons CQTC appeals to conserved quantities is to make a clear distinction between, on the one hand, objects such as shadows and spots of light which do not seem to be capable of undergoing causal interactions, and, on the other hand, causally 'potent' objects such as billiard balls, atoms or planets. A standard way for tracing such a distinction is via conserved quantities: in contrast to planets or billiard balls, shadows do not 'posses' quantities that are governed by conservation laws (even though they do seem to have properties such as shape or velocity, these are not governed by conservation laws). Once this is clear, it should be evident that almost-conserved quantities of the kinds proposed by Harte ([2008]) do not threaten a collapse of the distinction between shadows and billiard balls, a collapse that would undermine the very foundations of CQTC. The reason is that in GR, these almost-conserved quantities are predicated of the exact same kinds of objects (i.e., particles, planets, black holes, etc.) as the kinds of objects of which exact conserved quantities are predicated. And, importantly, the (almost-)conservation of these quantities is entailed by well-defined, law-like equations which follow from the formalism of GR (namely, the equations written by Harte ([2008], p. 15)), as opposed to, say, the velocities or the shapes of shadows.

In short, we do not see any significant tension between CQTC and the Almost Killing view that we have been discussing in this section; the inclusion of approximate symmetries (of the kind just discussed) and almost-conserved quantities does not seem to create any serious problems for CQTC. On the contrary, we think that if a slightly more relaxed understanding of conservation laws and conserved quantities is motivated by physical developments, then that is a good reason for adopting such understanding given that, as Dowe himself stresses, scientific practice is one of the main motivations in the development of CQTC (([2000], pp. 1–13)). It then seems to follow that a view that combines CQTC and AKF entails that CHALLENGE is false, for such a view recognizes that our spacetime can have causal processes and interactions even in the absence of (exact) Killing fields. Having said this, we want to strengthen the response to CHALLENGE by showing that there is another sense in which approximate symmetries can cast doubts on premise 4.

4.2 Asymptotic Symmetry

We said earlier that there are at least two main ways of making sense of the idea of approximate symmetries (and so two ways of rejecting premise 4). One is Almost Killing, which appeals to the notions of generalized affine collineations and Jacobi fields just discussed. The other one, which we will discuss now, is Asymptotic Symmetry, which appeals to the notion of asymptotically flat spacetimes (this second option is more widely discussed in the physics literature). In particular, it is well-known that asymptotically flat spacetimes allow for the definition of integral conservation laws, such as the laws corresponding to the ADM energy and the ADM linear momentum (e.g., see Jaramillo and Gourgoulhon ([2011], pp. 13–4) for details). These conserved quantities, found through the Hamiltonian formulation of GR, are the generators of asymptotic symmetries (roughly, symmetries that preserve the flatness conditions as we move to spatial infinity). For instance, invariance under asymptotic time translations is related to the conservation of ADM energy and invariance under asymptotic spatial displacements is related to the conservation of ADM linear momentum.[9]

[9]For our purposes, it is not important to discuss the various ways of making sense of asymptotic flatness. It suffices to say that there are coordinate-based approaches (where the coordinates at spatial infinity are taken to be

Now, our universe is (we assume) not asymptotically flat—however, when using GR (and scientific theories in general) in concrete applications, it is standard physics practice to focus on subsystems that are approximately isolated (that is, subsystems for which the effects of other subsystems can be neglected for theoretical or practical calculations). This means that these (effectively) isolated subsystems can be treated *as if* they were embedded in spacetimes that are asymptotically flat, and hence embedded in spacetimes that exhibit asymptotic symmetries and the conserved quantities associated with these symmetries. For example, when using GR to describe the behaviour of stationary black holes, rotating black holes, binary star systems, or a planet around a star, one assumes in all cases that these objects (or pair of objects when dealing with two bodies) are properly isolated from external influences, and so are modelled as if they were embedded in an asymptotically flat spacetime. Similarly, when using GR to model gravitational waves, it is assumed that the waves are well-defined (mathematically speaking) only if we go sufficiently far away from the source—that is, if we go into to regions where the spacetime approximates that of Minkowski (see, e.g., (D'Ambrosio *et al.* [2022])). Not only has the practice of modelling subsystems in GR as if they were embedded in asymptotically flat spacetimes been central for applying GR to actual systems in our universe, but, precisely because of that same fact, such practice has been central to the empirical confirmation of the theory. (This, indeed, has been true since the genesis of general relativity: see (Lehmkuhl [2017]).) For instance, when scientists say that the recent detection of gravitational waves has further confirmed GR, they say so in the context of taking seriously our models of gravitational waves, all of which model such detection as if the detectors (located here on Earth) were 'located' in regions of spacetime that are very far away from the source and nearly flat. Harte ([2008], p. 4), for example, says that 'The assumption of a simple limiting form for the geometry [that the geometry is asymptotically flat] makes it convenient to invariantly describe certain properties of a spacetime in terms of 'measurements at infinity' [...] Measurements like those expected from gravitational wave detectors do come very close to fitting into this formalism.' In short, then, to take seriously the success of GR requires one to take seriously the practice of modelling approximately isolated subsystems via asymptotically flat spacetimes. And that practice allows us to, *inter alia*, define various conserved quantities for the systems at hand and, even more importantly, provides us with a means of testing GR itself.

Asymptotic Symmetry counters premise 4 in CHALLENGE because even if our universe is not asymptotically flat and even if it lacks exact symmetries, one can still say that it approximately has the symmetries required for a derivation of integral conservation laws, in the following sense: our universe is such that it has approximately isolated subsystems, and such subsystems exhibit approximate asymptotic symmetries, in the sense that each can be adequately modelled as if it were within a spacetime that is asymptotically flat (for more on the emergence of symmetries for isolated subsystems, see (Wallace [2020])). Hence, we can define

Euclidean to first order), and there are geometric-based approaches that appeal to the conformal compactification picture. For a good introduction for these various approaches, see (Jaramillo and Gourgoulhon [2011]).

integral conservation laws for particular (effectively isolated) subsystems of the universe even if, for the whole universe, we can't do so (unless we were to model the whole universe as being asymptotically flat). And, even if these integral laws are in some sense derived from approximations/idealisations, the fact that physicists take these models seriously (e.g., take them as providing ways for testing GR) suggests the fact that the laws are, if not exactly, to the very least approximately associated to conserved properties of these subsystems. Now, before we show how CQTC can take advantage of Asymtptotic Symmetry, this is a good moment to consider some recent criticism by Duerr ([2019]) against taking seriously conclusions derived from the assumption of asymptotic flatness. He says that

> in no realistic scenarios does asymptotic flatness hold any longer: Our ΛCDM-universe isn't asymptotically flat, not even approximately. [...] Asymptotic flatness should thus be viewed as an idealisation in the sense of Norton (2011): Within a certain regime, it adequately models some aspects of subsystems of our universe by dint of a distinct surrogate system which approximately mimics distinctive features of the target system. But with that, the predictive and explanatory success of asymptotically flat idealisations no longer warrants an unqualified realism about all explanantia involved—in particular, about [gravitational] energy. ([2019], pp. 29–30)

Essentially, Duerr is worried about inferring that quantities such as gravitational energy are real when such quantities are derived from asymptotically flat spacetimes, and he is worried about this because asymptotically flat spacetimes are idealizations; as he points out, our universe is not asymptotically flat, but rather asymptotically de Sitter. So, if Duerr is correct, then it seems that we cannot really take very seriously Asymtptotic Symmetry, as it makes essential use of the symmetries of asymptotically flat spacetimes in order to construct integral conservation laws for various quantities (including ADM energy).

Having said this, we find Duerr's criticism to be unconvincing for at least two reasons. First—and more straightforwardly—it seems to neglect the fact that there are well-known ways of deriving integral conservation laws for asymptotically de Sitter spacetimes (e.g., (Abbott and Deser [1982]) or (Szabados and Tod [2019])). So if Duerr is worried about the fact that our universe is asymptotically de Sitter, then the fact that asymptotically de Sitter spacetimes also allow for the construction of integral conservation laws should alleviate a good part of his worries.

Second—and this is much more important—we think that Duerr's criticism is in tension with the way models of approximately isolated subsystem are intended to be used and interpreted by physicists. When physicists use, say, the Kerr solution to model a black hole, they know that not every bit of the mathematical model ought to be taken literally, and this is in particular true of the fact that the infinite limit (when r goes to spatial infinity) in the model cannot be interpreted as if we were approaching a physical distance that is really infinitely away from the black hole.

However, it is still possible to take seriously (as opposed to as a mathematical artifact of the model) some properties defined via the models even if these models are idealizations in various ways. For a central idea when modeling subsystems in physics is that the properties that we can derive from them, when they are assumed to be perfectly isolated in a particular given model, resemble, to a good approximation, some of their actual properties in those situations in which the actual (target) system is approximately isolated. An example can be useful. Imagine a very simple universe with just one black hole in it. That is a simple case of an asymptotically flat spacetime. Then, an asymptotically flat model of that universe provides a very good fit of the actual system, and we should not be worried about the asymptotic flatness in this model (in this case, the asymptotic flatness would be a feature of the universe too). In such a case, we should not worry about taking the ADM mass or the ADM momentum of the black hole system (the spacetime in which the black hole is embedded is part of the system, more on this soon) to correspond to some sort of real physical property of the black hole system. Now consider a case of a simple universe but, instead of one, now there are two black holes that are extremely far away from one another (so far that in the middle point between them, no experiment—given technological limitations—could detect the slightest presence of curvature). Then, suppose that in such a case, we model each individual black hole via a model like the one used before, namely, a model of an asymptotically flat spacetime with just one black hole in its 'center'. Question: should we no longer take seriously the ADM energy or ADM momentum of either black hole system on the basis that such quantities were derived from models that only consider one black hole as opposed to two? Not at all! We think that if earlier it made sense to attribute to the black hole system an ADM energy E, now it should make sense to attribute that same energy to each individual black hole system as long as they are appropriately isolated from one another. This simple example highlights an important feature of more realistic scenarios, namely, that the properties that we can derive from models of isolated subsystems (e.g. ADM energy) are physically interesting precisely because they can be taken as reflecting certain properties of actual subsystems of the universe insofar as they are appropriately isolated from one another (otherwise, there would not be much of a point in theorizing about isolated subsystems to begin with!).[10]

In short, we think that it is rather premature to say that we can't take those properties of actual systems in GR that are derived via models that are asymptotically flat as being genuine properties of the systems. And, in any case, we do not need to think that the integral conservation law that we write for these quantities corresponds in an exact way to some physical property of the system. It would suffice that the physical system has a quantity that is very well represented by such a conservation law even if, perhaps, such a quantity is not exactly conserved (as we discussed for the case of Almost Killing above). Skepticism about the reality of any quantity

[10]One might seek to defend Duerr here as follows: 'unqualified realism is not warranted' is not the same as 'we should no longer take ADM energy or momentum seriously'. We agree—but we maintain that Duerr does indeed take the stronger line, as is evident when one reads his article in its entirety.

that is derived via a model that makes essential use of the asymptotic flatness condition would lead to a rather generalized skepticism with respect to the application and understanding of GR. (This being said, of course one will ultimately need to be precise and concrete about exactly when asymptotically flat models can be applied—we do not address this question here, interesting and important as it is.) Indeed, this kind of skepticism would amount to some sort of skepticism about GR itself, as these asymptotically flat spacetimes have played a central role in the theoretical and practical developments of the theory.[11]

Say, then, that we adopt the Asymptotic Symmetry strategy, and so we take seriously certain conserved quantities of (approximately) isolated subsystems such as the ADM energy. In that case, we still want to consider whether or not such a strategy is indeed compatible with CQTC. As far as we can tell, the answer is not very clear. On the one hand, saying that isolated subsystems have conserved properties such as the ADM energy is certainly a good thing from the perspective of CQTC, for we want to say that (approximately) isolated subsystems such as black holes are causal objects (i.e., objects capable of entering in causal relationships) and, in order to say so in the framework of CQTC, we need to ascribe to them properties that can be governed by conservation laws. On the other hand, notice that these kinds of properties are not really properties of just the subsystem in question (e.g., an isolated black hole) but rather properties of the subsystem plus the asymptotically flat spacetime in which it is assumed to be embedded. To be specific, note that the ADM energy really represents the total energy in a particular foliation of spacetime, not just the energy of a particular localized subsystem in such a foliation (Jaramillo and Gourgoulhon [2011], p. 14). This means that, technically speaking, if we were to adopt CQTC in this context, we should not say that a given (isolated) black hole counts as a causal object *per se*, but rather, we should say that the black hole and a significant portion of the spacetime around it (a portion that becomes approximately flat far away from the black hole) counts as a causal object. So it is not at all clear that Asymtpotic Symmetry together with CQTC recovers the kind of intuitive picture that we might have desired, namely, one where a given subsystem such as a black hole is treated as a causal object that can causally interact with other systems such as a planet close by. Rather, as far as CQTC is concerned, it seems that Asymptotic Symmetry requires that we treat the black hole and a significant spacetime region 'around' it as a *single* causal object in such a situation. In any case, even if CQTC together with Asymptotic Symmetry do not recover a natural picture with respect to what counts as a causal object, we want to stress that they entail that there *are* causal objects nonetheless (e.g., a black hole together with a significant spacetime region 'around' it). Hence, the conjunction of CQTC and Asymptotic Symmetry undermines CHALLENGE. Granted, simply showing counterexamples to CHALLENGE is not sufficient to motivate that CQTC and Asymptotic

[11]One might wonder how asymptotic flatness interacts with 'GR principles' such as 'background independence.' For an illuminating investigation into these issues, see (Belot [2011]). But whatever the outcome of these investigations, it would quite transparently be a mistake to use said principles to rule out the application of asymptotically flat spacetimes, given the ubiqity of the latter. (For more on background independence, see (Pooley [2017]) and (Read [2016]).)

Symmetry work well together, and it would be interesting in future investigations to explore how much the clash between the way in which we usually individuate astrophysical objects (e.g., 'there is a black hole in the center of the galaxy') and the way Asymptotic Symmetry does so leads to deeper conceptual issues as regards causation, laws and natural properties.

Before we move on to consider other ways of responding to CHALLENGE, a quick recapitulation might be helpful. The main upshot of the previous two subsections is that whether or not our universe admits of symmetries, and whether or not it admits integral conservation laws, is not as simple as just having exact Killing fields (despite what premise 4 suggests). As Almost Killing highlights, there are approaches in the physics literature for defining symmetries and integral conservation laws even in contexts of spacetime lacking exact Killing fields. And as Asymptotic Symmetry shows, even if the universe as a whole lacks (exact) Killing fields, it is still possible to ascribe symmetries and integral conservation laws to approximately isolated subsystems. Taken together, these two approaches constitute good evidence that, at the very least, premise 4 needs much stronger justification than what has been provided in the literature so far.

5 Defences of local conservation

In the previous section, our strategy for countering CHALLENGE consisted in questioning premise 4 on the grounds that whether or not our spacetime admits of symmetries is a complex matter not settled simply by its having or lacking exact Killing fields (we also questioned premise 2, as we showed that one can define integral conservation laws from approximate symmetries). This, by itself, shows that the standard criticism of CQTC in the context of GR (as represented in CHALLENGE) is not satisfactory and needs to be revised. But now we will reinforce the response by shifting gears and attacking premise 3, which states that 'kind 2' quantities (those which admit of differential but not of integral conservation laws) are pathological. In Section 5.1, we will examine arguments for the claim that local conservation laws are pathological, and then we will present some responses. Then, in Section 5.2, we will consider the following question: even if we assume that local conservation laws are not problematic *per se*, can such laws be brought to the defense of CQTC? Here, we will consider three reasons why such laws do not really suffice within CQTC's framework, and then we will consider various approaches to solving these problems.

5.1 Problems with $t^{\mu\nu}$

To start, let's consider two possible versions of differential (as opposed to integral) conservation laws for energy in the context of GR. As these laws are local, they would obtain even in spacetimes without Killing vector fields. As such, if they were to be deemed acceptable, physically significant conservation laws, they would block CHALLENGE for they would afford a means by which we could appeal to conservation laws in any spacetime, including our own (and, the thought is, once we have 'access' to these laws, we can apply CQTC).

The first candidate for a local conservation law is the vanishing of the covariant divergence of the stress-energy tensor, $\nabla_\mu T^{\mu\nu} = 0$. This follows from the Einstein equation and the Bianchi identities; however, the interpretation of this equation as a conservation law is not straightforward, as it is not a partial divergence (Hoefer [2000]).[12] That said, essentially since the genesis of GR (with the work of Noether, Hilbert, Einstein, and Klein), a candidate 'partial divergence' conservation law has been discussed extensively, namely, $\partial_\mu (T^{\mu\nu} + t^{\mu\nu}) = 0$.[13] This expression can actually be derived via application of Noether's theorems to GR (see (Trautman [1962]), and for a more contemporary discussion, (De Haro [2022])). Here, the pseudotensor $t^{\mu\nu}$ would correspond to gravitational stress-energy, and the tensor $T^{\mu\nu}$ to the energy associated with matter fields (as usual). (Note that $t^{\mu\nu}$ is not necessarily unique, for it is defined only up to a superpotential factor: see (Trautman [1962]).) Notice, in particular, that this expression suggests a natural interpretation whereby the energy associated with matter fields is balanced out by the energy associated with gravity (see (Lehmkuhl [2008]) for discussion of what represents the gravitational field in GR).

The literature has focused on two problems with local conservation laws of the type just discussed: (a) pseudotensorial objects such as $t^{\mu\nu}$ are not 'geometric objects', which is to say that they do not have well-defined transformation rules relating their coordinate representations.[14] This appears to preclude the possibility of ascertaining the coordinate-independent reality associated with said coordinate representations (via appeal to, say, the Kleinian approach to geometry—see (Wallace [2019])—which proceeds by identifying the invariants associated with such representations and the transformations between them), and thereby to raise the threat that, if such objects were to be taken seriously, reality would be fundamentally perspectival (for further discussion of this issue, see (Read [2022])).

Moreover, as already indicated in parentheses above, (b) there are infinitely many such pseudotensors, and so (it would appear) infinitely many conservation laws in GR. There being infinitely many such conservation laws appears *prima facie* to be problematic because conservation laws in physical theories should (the thought might go) be very special kinds of laws, usually hard to come by and scientifically interesting (they are explanatorily powerful, have predictive power, and so on), and yet their apparent infinite number in GR seems to suggest that they are actually trivial (in the sense of being ubiquitous and easy to come by).

These first two problems with local stress-energy conservation in GR (namely, the non-geometrical character of $t^{\mu\nu}$ and the abundance of these conservation laws) have been recognized in the literature before (see e.g. (Hoefer [2000]) and (Lam [2011])). Indeed, it would be natural to say that the received wisdom is that these two problems are sufficiently significant

[12]It is also worth noting how weak this condition really is: see (Pitts [2022b]) for discussion.

[13]See e.g. (De Haro [2022]) and references therein for both technical and historical background.

[14]For a clear presentation of the sense in which this is so, along with references to the historical literature on geometric objects, see (Duerr [2021]).

that differential (local) conservation laws in GR are just too pathological to be taken seriously.

Having said this, we want to point out that in opposition to the received wisdom, some authors (in particular Pitts ([2022a]), Read ([2020]), and De Haro ([2022])) have argued that (a) and (b) are not problematic. Regarding (a), why should geometric objects be the *sine qua non* of physical theorising? On the one hand, one could embrace the perspectivalism which appears to go hand-in-hand with the use of such objects (Read [2022]). On the other hand—and more conventionally—one could point out that non-geometric objects abound in physics (most notably, consider the case of spinors), and thereby ask: if one is willing to be a realist in the case of those other quantities and objects, why not be a realist here also? Regarding (b), one should (the above authors—in particular Pitts ([2010])—claim) trust the mathematics of Noether's theorems: if these theorems state that there are infinitely many gravitational energies in general relativity, then so be it. Incidentally, Pitts ([2010]) claims that, in light of there being infinitely many such laws, general relativity enforces conservation *more* strongly than other theories.[15]

For the reasons just exposed, we do not find (a) and (b) to be particularly damning problems for the defender of local conservation laws in GR, and, consequently, we do not find premise 3 in CHALLENGE to be particularly well-motivated. However, even if we were to take seriously kind 2 quantities, the defender of CQTC would still have to show that local conservation laws can still do the work required by CQTC. Investigating this matter is the focus of the next section.

5.2 Are $t^{\mu\nu}$ and CQTC compatible?

Even settling the above apparent problems regarding $t^{\mu\nu}$ and its associated conservation laws, one might have concerns regarding whether such an object can be appealed to by proponents of CQTC as they attempt to defend their approach to causation in the context of GR. In this subsection, we articulate and address such concerns.

5.2.1 Three reasons for the incompatibility of $t^{\mu\nu}$ and CQTC

In the literature exploring the compatibility between GR and CQTC, one can identify two alleged problems with using local conservation laws together with CQTC. One is a problem highlighted by Curiel ([2000]) having to do with the idea that the notion of 'exchange' that is crucial for CQTC is not adequately captured by local conservation laws but only by integral laws. He says, in particular, this:

> The idea that some stuff is being transferred or is propagating refers of necessity to more than one spacetime point, that or those from which and that or those to which, and whereas the differential form of the conservation principle refers only to one spacetime point, it is the integral form that represents the fact that the stuff is conserved over finite spatiotemporal volumes. (Curiel [2000], p. 45)

[15] For further responses to (b), see (Pitts [2022c], §10).

A second problem—this time suggested in the writings of Lam—states that the fact that gravitational energy is not represented by specifically a quantity invariant under coordinate transformations suggests that such a quantity is 'non-local' in a troubling sense, because its components can be transformed to zero by way of a particular choice of coordinate system. (Note—which is arguably not completely transparent in Lam—that this concern is distinct from $t^{\mu\nu}$ not being a geometric object because some *geometric* objects, such as connection coefficients, can also have their components transformed to zero by suitable coordinate choices.) In particular, he writes:

> It is always possible at any spacetime point to find a coordinate system in which (infinitesimally) there is no gravitational energy. As a consequence, gravitational energy can be understood as non-local in the precise sense that the amount of gravitational energy in any given spacetime region cannot be defined in a unambiguous way [...] The characterization of "causal" for an interaction then loses its fundamental meaning and one cannot say unambiguously whether two events are causally related or not (Lam [2010], pp. 66–7)

Lam's concern here appears to be that given that one can always find a coordinate system in which the components of pseudotensorial quantities such as $t^{\mu\nu}$ vanish, whether or not a given interaction in GR is causal or not would depend, according to CQTC, on the coordinates used. If the coordinates are such that the conserved quantity of an object vanishes, then it would follow, according to Lam, that such an object would not be able to 'participate' in causal interactions.

In addition to these two problems (the 'transfer' issue raised by Curiel and the issue just mentioned by Lam), we actually think that there is a third potential problem regarding the compatibility of CQTC and GR. To see why, notice that it is natural to expect that for a given GR phenomenon that is apparently causal, say a gravitational wave heating a moon of Jupiter (on which more below), a good theory of causation should be able to support a somewhat unified story: the gravitational wave did this and that, and then exchanged energy with the moon according to a certain law, and then the moon was heated by a given amount, etc. However, if we take both $\partial_\mu (T^{\mu\nu} + t^{\mu\nu}) = 0$ *and* CQTC seriously, then what we really (at least appear to) have are infinitely many causal stories of the apparently same unified phenomenon: depending on the particular superpotenial used, we might pick out a different particular conservation law out of this family when 'reconstructing' the causal story. Hence, instead of a single story, what we seem to have according to CQTC is a multitude of 'parallel' stories, each appealing to a different type of conserved quantity and hence to a different type of causal process and causal interaction.

5.2.2 Solving the three problems of incompatibility

We will now address the previous problems, starting with Curiel's claim that local or differential conservation laws cannot adequately represent the notion of 'exchange' or 'transfer'. It is true, as Curiel suggests, that the notion of a quantity being transferred/exchanged is crucial to physical connection theories of causation. However, we do not see why such a notion cannot be captured by the differential version of a conservation law, for two reasons. First, the differential version of the law is not really about just one spacetime point, as Curiel implies, but rather about an infinitesimal spacetime region (see, e.g., (Butterfield [2006])). Second, we do not agree with Curiel when he suggests that the only way of making sense of 'transfer' is via the integral form of the law. In electrodynamics, for example, it is very natural to represent the *transfer* of charge q via the continuity equation $\nabla \cdot \mathbf{J} + \frac{\partial \rho}{\partial t} = \sigma$, where σ is the source for q, $\mathbf{J}$ is the flux of q, and ρ is the amount of the quantity q per unit volume.

Curiel might respond that a differential conservation law is surely less operationally significant than an integral conservation law, since physical measurements, experiments, and observations occur over finite regions.[16] We agree with this—but it is not clear (to us at least) why such operational considerations should be relevant to the *metaphysical* question as to whether causation exists according to CQTC; for this, we maintain that a differential conservation law appears perfectly sufficient.

Let's move on to Lam's concern that one can always find a coordinate system in which the components of pseudotensorial quantities such as $t^{\mu\nu}$ vanish. It is not completely clear whether Lam is saying that the fact that one can find such a coordinate system for these objects is *per se* problematic, or saying that such a fact is problematic in the specific context of taking CQTC and GR jointly (e.g., Lam might be worried by the fact that if a quantity vanishes in some coordinate systems and not in others, then it seems to follow from CQTC that a given process would not be causal in certain coordinate systems and causal in others). Either way, we think Lam's concerns can be addressed. If Lam means the former concern, we note that one can also find such coordinate systems for (say) connection coefficients in general relativity, or (more prosaically) for kinetic energy in classical mechanics. In both the connection coefficient case and the kinetic energy case, there seems to be little reason to regard such mathematical objects (or the quantities they are taken to represent) as being 'non-local' in any problematic sense, in which case, one might ask why one cannot say the same thing about objects such as $t^{\mu\nu}$.

If Lam means the second kind of concern, it is useful to recall that CQTC does not require that a given process has a non-zero conserved quantity in order to count as a causal process. Rather, CQTC requires that the object instantiates the quantity in question, even if the *value* of such quantity (in a given coordinate system) is zero. To better see this, note that it is uncontroversial that CQTC applies to simple cases of billiard balls colliding with one another, and these cases involve objects possessing quantities whose values are clearly frame-dependent such as

[16] Indeed, Curiel has suggested this line of response to us in personal communication.

kinetic energy or linear momentum. Hence, neither the frame-dependence of (the values of) a physical quantity nor the zero value of a physical quantity, by themselves, make these quantities unacceptable or unsuitable for their use in physical connection theories such as CQTC.

Finally, the apparent disunity in the description of causal stories (the third problem above), coming from the choice freedom associated to the superpotential, is not as problematic as it might seem at first. After all, we can regard these apparently different causal descriptions of the same phenomenon as just different ways of looking at the same phenomenon from different perspectives, and, importantly, different perspectives that agree about the most important features of the 'causal story.' That is, one could say that what really matters when providing a causal description of the phenomenon within the framework of CQTC is not so much what particular conserved quantity we focus on, but the fact that the object in question has a conserved quantity, and the fact that such quantity (whichever we pick) is conserved and it is exchanged during the interaction with other objects. And if one thinks this, then the different descriptions (parametrized by different choices of the conserved quantity coming from different choices of superpotentials) agree on these basic points.[17] An analogy might help: when describing the elastic collision of two billiard balls within the framework of CQTC, one can use a description that focuses on the exchange of linear momentum, or one can use a description that focuses on the exchange of kinetic energy (both descriptions agree in that each billiard ball corresponds to a causal process, and that when the balls meet there is a causal interaction).

Now, there might still be a worry about possible differences in concrete applications when using different quantities (or different pseudotensorial laws) for gravitational energy. To be more precise, consider the following question: what guarantees that the amount of gravitational energy absorbed by a given body subjected to a strong gravitational field does not depend on whether we use Einstein's pseudotensor instead of, say, Landau-Lifshitz' pseudotensor? Thankfully, there are some indications that, at least in certain important applications, the choice of pseudotensor does not matter. In particular, Favata ([2001]), who was worried precisely about this kind of question, showed that when calculating the tidal work that a gravitational field does on an isolated body (e.g., a moon of Jupiter), such work is exactly the same whether one uses Einstein's, Landau-Lifshitz' or Møller's pseudotensors (such work is given by equation (1) in ([2001])). Hence, in concrete terms, if one calculates the heating of Io due to Jupiter's gravitational field, one obtains the same result independently of the particular choice of pseudotensor. For Favata ([2001], p. 2), this is encouraging because 'it is clear that the tidal work is a physical observable and should in no way depend on one's means of calculating it.' For us, this is at least a reason to alleviate the worry that the use of different pseudoten-

[17] Jacobs ([2021]) makes a similar point in the case of different choices of the gauge potential in the Aharonov-Bohm effect. Note that the relation between the different superpotentials isn't straightforward: in our view, one illuminating way of understanding this relationship is that different superpotentials correspond to different pullbacks along local sections of the 'Nester-Witten form', which is an object defined on the bundle of linear frames. (Clearly, we can't go into this further here, but see (Szabados [1992]) for the relevant physics and mathematics, and (Jacobs and Read [2022]) for a discussion of the extent to which this way of thinking helps to overcome the problem of the non-uniqueness of pseudotensors.)

sors might lead to disunity in the causal story (in this case, such disunity would amount to a disagreement on the amount of gravitational energy performing work).

It is worth pointing out that the treatment of Io's gravitational heating just described applies to any other case in which a body is in the presence of an external gravitational field whose 'external curvature is nearly uniform, and the spacetime curvature is not changing too rapidly' ([2001], p. 8).[18] Examples of such systems include cases of small bodies orbiting more massive objects, such as Io's case or the case of the moon around the Earth, but can also include a neutron star or a black hole in a binary system that is not yet too close to merging ([2001], p. 7). It is also worth pointing out that the framework of Favata ([2001]) is not limited to calculating the heat generated by an external field, but can also be used to determine the deformation or the vibrations induced by said field on a body. It then seems that at least for billions of concrete applications of two bodies gravitating, the results of computations related to the tidal work do not seem to depend on the choice of pseudotensor.[19] And this is obviously good news for those interested in using pseudotensors in the framework of CQTC.

The work by Favata ([2001]) is also relevant to us in that it seems to adopt a methodology very close in spirit to CQTC. To see this, let us summarise the main steps in the calculations presented in that article. Favata begins by writing $\partial_\mu (T^{\mu\nu} + t^{\mu\nu}) = 0$, and points out that this conservation law shows that massive objects can exchange energy with the gravitational field ([2001], p. 4). If one adopts CQTC, then one could say that an object with energy $T^{\mu\nu}$ together with a gravitational field around it with energy $t^{\mu\nu}$ constitute, when taken as a whole, a causal process. Second, Favata ([2001]) uses $t^{\mu\nu}$ in order to compute expressions for a conserved quantity (e.g., the four-momentum) for the total system (henceforth 'SYSTEM'), in this case, the system given by Io together with a region (the 'buffer region') around it ([2001], p. 5). Third, Favata ([2001]) identifies the rate of change of the first component of the four-momentum computed earlier to consist of both the interaction energy (the energy associated with the interaction between the body's quadrupole momentum and the external field) and the tidal work (the work the external field exerts on the body). Here, it is crucial not to confuse the external gravitational field that performs the tidal work, and the gravitational field associated with the SYSTEM. The former is assumed to be produced by an external body such as Jupiter, whereas the latter, associated originally with $t^{\mu\nu}$, is the field that characterizes the SYSTEM. Of course, one can associate $\partial_\mu (T^{\mu\nu} + t^{\mu\nu}) = 0$ with either field, and so one can treat either field as belonging to a causal process according to CQTC. Fourth, Favata ([2001]) shows that the expression for the tidal work does not depend on the particular pseudotensor used, whereas the expression for the interaction energy does (physically, this means that, unlike the energy attributed to the field, the energy transfer between the external field and the body is invariant under different choices of pseudotensors). Recall that in the framework of CQTC, changes in a conserved quantity are

[18]In slightly more technical jargon, this means that we adopt the *slow-motion* approximation, and we neglect the contributions of the target body on the field.

[19]It is an encouraging sign that quasi-local approaches to tidal work—which rely on tensorial quantities—also seem to give the same results (e.g., see (Booth and Creighton [2000]) or, more recently, (Huber [2022])).

associated with the presence of a causal interaction. This is precisely what is happening here, as the author takes the rate of change of the first component of the four-momentum in order to compute the tidal work, which (in Favata's own words) corresponds to the *energy transfer* between the external field and the SYSTEM ([2001], p. 2).

In summary, CQTC's framework seems to fit rather nicely with this project, particularly concerning the use of conserved quantities and their variation when modeling causal interactions.[20] In any case, we hope that the work of Favata ([2001]) illustrates both that the multiplicity of pseudotensors need not inevitably lead to the disunity of the causal story, and, perhaps more importantly, that physicists working in GR have modeled certain putative cases of causal interactions by means of changes in certain conserved quantities, very much in the spirit of CQTC.

6 Conclusions

The literature investigating the compatibility of CQTC and GR has focused on two main problems: (i) difficulties associated with defining integral conservation laws in GR, and (ii) difficulties associated with local conservation laws such as the one given by Einstein's pseudotensor. In this paper, we have shown how recent work in the foundations of GR can help a proponent of CQTC to address these two problems. In particular, regarding (i), we showed that even in the absence of exact Killing fields, there are methods for defining integral conservation laws (e.g., via almost-Killing fields, or via asymptotic symmetries). Regarding (ii), we showed that the main objections raised against local conservation laws (e.g, the non-geometric character of pseudotensorial conservation laws) do not really pose a serious obstacle for the defender of CQTC, and we also showed how some physicists have used certain pseudotensorial quantities in order to calculate interactions in GR in a way very much in the spirit of CQTC.

We want to point out, however, that there are still interesting challenges ahead for CQTC in the context of GR. To illustrate some of these challenges, it is helpful to consider a binary star system, consisting of two stars orbiting around their common center of mass. It is a standard line that the binary system causes gravitational waves to propagate. But now imagine that we want to use CQTC in order to make sense of this causal claim. The first step would be to identify three causal processes, namely, one for each star and then one for the waves (i.e., some particular region with waves). As should be clear from Sections 4 and 5, there are various ways of doing so (either via quantities of the whole spacetime such as the ADM energy or the Bondi energy, or locally via, e.g., Einstein's pseudotensor or using Harte's methods). The second step would be showing that whatever quantity we choose, it is such that it is 'exchanged' among the various systems (for if it is not, then we cannot say that there are causal interactions among

[20] This is not to say that there aren't open questions as to the extent that CQTC captures this kind of project. For instance, it is not completely clear from the calculations if the increase (decrease) in the four-momentum associated with SYSTEM can be understood as a decrease (increase) in the four-momentum associated with Jupiter's field, as condition D2 of CQTC would demand (one would expect this to be the case, as the energy is coming from Jupiter's field, but an actual proof would require carrying out a lengthy computation).

these systems). And here is where the troubles really begin, for the standard computations of the energy carried by gravitational waves assume that we are far away from the source (D'Ambrosio *et al.* [2022]). It is not clear that we can even properly talk of a well-defined wave near the source, and so it is not even clear that we can compute the energy of the 'later-to-become-wave' near the stars! Indeed, the standard approach to the energy of the waves appeals to the Bondi energy, which, as with the ADM energy, is a global property of spacetime (and not just a property of the binary system *per se*) that is evaluated at null infinity (the ADM energy is evaluated at spatial infinity). But the Bondi energy would not help us here, for CQTC requires an exchange of conserved quantities occurring at the intersection of the corresponding wordlines (in this case, the one associated with the waves and the one associated with the stars). Clearly, the waves reaching 'infinity' are not close to interacting with the stars that produced them. One might try to appeal to local approaches, such as Einstein's pseudotensor, but that would not solve the main issue, which is ultimately that the waves are not well-defined near the source, and so it is not clear how to even go about attributing energy to them (or to whatever would later become 'them') in those regions where presumably they causally interact with the stars. As this example illustrates, once we attempt to use the formalism of CQTC in concrete GR situations, problems start to arise.

Having said this, we believe that the difficulties faced by CQTC in concrete GR applications are actually symptoms reflecting both technical and conceptual difficulties of GR *per se*. In GR it is famously hard to model *prima facie* simple systems, such as the two-body problem. Indeed, the two-body problem is not in general solvable in GR, and it was actually less than 20 years ago when the first numerical techniques were developed capable of extracting detailed information about the evolution of binary systems close to merging and their production of gravitational waves (Pretorius [2005]). In such cases, where numerical methods are invoked, it is not really a surprise that CQTC cannot help much (at least at present), for there remain open questions regarding the physics involved.[21] In addition, there is a technical problem stemming from the fact that there are many approaches for defining energy in GR (quasilocal, global, etc.), and the extent to which these various approaches are related to one another is far from clear (for a detailed review, see (Szabados [2009])).

To end, let's consider a possible worry: do the conceptual difficulties that CQTC seems to face in the context of GR imply that other causal theories are preferable? The answer is not so simple. As Curiel ([2015]) and Jaramillo and Lam ([2021]) argue (see also (Vassallo [2019])), coming up with a good way of evaluating counterfactual statements in GR is a rather tricky issue in part because it is not clear how to make sense of nearby possibilities when comparing different spacetimes, in part because of the complexity of the initial value problem in GR. In

[21]At the very least, there is more to be said regarding how numerical methods and simulations can deliver understanding of the physics involved: see (Patton and Curiel [2023]) for recent engagement with these issues in the context of gravitational wave astrophysics. To stress: this is a different problem to those typically raised for CQTC (by e.g. Curiel ([2000])) because it is a problem of our limitations in physical modelling, rather than a philosophical problem having to do with CQTC *per se*.

addition to these difficulties, some might argue that counterfactual theories ‘produce’ more causation than what one might expect in the context of GR. To see the thought behind this, consider this claim: if the Earth had not been located where it was five minutes ago, then a given asteroid would not have accelerated in the way it did. Assuming a counterfactual theory, one might take that claim to support that the Earth caused the asteroid to accelerate in the way it did. Furthermore, one might think that it is a virtue of a counterfactual theory that, in contrast with CQTC, it captures apparently straightforward causal situations like this one even in the absence of a full story about the physics of the situation (i.e., even in the absence of a solution to the two-body problem in GR). However, things get tricky once we realize that, technically speaking, the asteroid was always in free fall (i.e., it was just following a geodesic). If an object is following a geodesic, then it just seems wrong to say that something else (say the Earth) is causing it to do so—after all, geodesics are what all objects are supposed to follow in the absence of external factors. This is like saying that something caused an isolated object in Newtonian mechanics to keep moving inertially. This is why, arguably, counterfactual theories might lead to more causation than what is warranted by GR itself. To be clear, we offer this as one example of the kinds of conceptual issues which counterfactual theories of causation face in the context of GR, rather than as a knock-down argument against those theories in that context. For better or for worse, our causal intuitions regarding causal interactions in space (moons orbiting planets, spaceships using the Earth’s gravity to accelerate, etc.) assume Newtonian gravitation, where counterfactual analyses seem to work relatively well. But very much as in with the transition from Newtonian mechanics to quantum mechanics, the transition from Newtonian gravity to GR requires a deep revision of our intuitions, including those related to causal interactions. Perhaps what this all suggests is that causation is a much subtler matter in GR than what we might have expected.

Funding

This work was supported by grant number 61521 from the John Templeton Foundation.

Acknowledgements

We are very grateful to Erik Curiel, Henrique Gomes, Abraham Harte, Caspar Jacobs, and Brian Pitts for helpful and encouraging feedback. We also are appreciative of the comments that we received from the referees. Sebastián thanks the philosophy community at the University of Oxford for their support during his postdoc, and Andrés would like to extend a note of thanks to the late Paul Nuñez.

Sebastián Murgueitio Ramírez
Philosophy Department
Purdue University

United States
smurguei@purdue.edu

James Read
Faculty of Philosophy
University of Oxford
United Kingdom
james.read@philosophy.ox.ac.uk

Andrés Páez
Department of Philosophy
Universidad de los Andes
Colombia
apaez@uniandes.edu.co

References

Abbott, L. and Deser, S. [1982]: 'Stability of gravity with a cosmological constant', *Nuclear Physics B*, **195**(1), pp. 76–96. <https://www.sciencedirect.com/science/article/pii/0550321382900499>

Belot, G. [2011]: 'Background-Independence', *General Relativity and Gravitation*, **43**, pp. 2865–2884.

Booth, I. S. and Creighton, J. D. E. [2000]: 'Quasilocal calculation of tidal heating', *Physical Review D*, **62**(6), pp. 1–4.

Butterfield, J. [2006]: 'Against Pointillisme about Mechanics', *The British journal for the philosophy of science*, **57**(4), pp. 709–753.

Curiel, E. [2000]: 'THE CONSTRAINTS GENERAL RELATIVITY PLACES ON PHYSICALIST ACCOUNTS OF CAUSALITY', *Theoria (Madrid, Spain)*, **15**(1(37)), pp. 33–58.

Curiel, E. [2015]: 'If Metrical Structure Were Not Dynamical, Counterfactuals in General Relativity Would Be Easy', *arXiv preprint arXiv:1509.03866*.

D'Ambrosio, F., Fell, S. D. B., Heisenberg, L., Maibach, D., Zentarra, S. and Zosso, J. [2022]: 'Gravitational Waves in Full, Non-Linear General Relativity', .

De Haro, S. [2022]: 'Noether's Theorems and Energy in General Relativity', in J. Read and N. Teh (*eds*), *The Physics and Philosophy of Noether's Theorems*, Cambridge University Press.

Dowe, P. [2000]: *Physical Causation*, Cambridge studies in probability, induction, and decision theory. Cambridge University Press.

Duerr, P. M. [2019]: ‘Fantastic Beasts and where (not) to find them: Local gravitational energy and energy conservation in general relativity’, *Studies in History and Philosophy of Modern Physics*, **65**, pp. 1–14.

Duerr, P. M. [2021]: ‘Against ‘functional gravitational energy’: a critical note on functionalism, selective realism, and geometric objects and gravitational energy’, *Synthese*, **199**(2), pp. 299–333.
<https://doi.org/10.1007/s11229-019-02503-3>

Favata, M. [2001]: ‘Energy localization invariance of tidal work in general relativity’, *Phys. Rev. D*, **63**, pp. 064013.
<https://link.aps.org/doi/10.1103/PhysRevD.63.064013>

Fletcher, S. C. [2020]: ‘Approximate local Poincaré spacetime symmetry in general relativity’, in T. S. Claus Beisbart and C. Wüthrich (*eds*), *Thinking about space and time*, pp. 247–267.

Fletcher, S. C. and Weatherall, J. O. [2022]: ‘The Local Validity of Special Relativity, Part 1: Geometry’, *arXiv preprint arXiv:2206.10768*.

Harte, A. I. [2008]: ‘Approximate spacetime symmetries and conservation laws’, *Classical and quantum gravity*, **25**(20), pp. 205008.

Hoefer, C. [2000]: ‘Energy Conservation in GTR’, *Studies in History and Philosophy of Modern Physics*, **31**(2), pp. 187–199.

Huber, A. [2022]: ‘Remark on the quasilocal calculation of tidal heating: Energy transfer through the quasilocal surface’, *Physical review. D*, **105**(2).

Jacobs, C. [2021]: *Symmetries as a Guide to the Structure of Physical Quantities*, University of Oxford, D.Phil. thesis.

Jacobs, C. and Read, J. [2022]: ‘Absolute representations and modern physics’, Unpublished.

Jaramillo, J. L. and Gourgoulhon, E. [2011]: *Mass and Angular Momentum in General Relativity*,, pp. 87–124 Dordrecht: Springer Netherlands.
<https://doi.org/10.1007/978-90-481-3015-3$_4$ >

Jaramillo, J. L. and Lam, V. [2021]: ‘Counterfactuals in the Initial Value Formulation of General Relativity’, *The British journal for the philosophy of science*, **72**(4), pp. 1111–1128.

Lam, V. [2010]: ‘Metaphysics of causation and physics of general relativity’, *HUMANA. MENTE Journal of Philosophical Studies*, **4**(13), pp. 61–80.

Lam, V. [2011]: ‘Gravitational and Nongravitational Energy: The Need for Background Structures’, *Philosophy of Science*, **78**(5), pp. 1012–1023.

Lehmkuhl, D. [2008]: 'Is spacetime a gravitational field?', in D. Dieks (*ed.*), *The Ontology of Spacetime II*, Elsevier.

Lehmkuhl, D. [2017]: 'Literal versus Careful Interpretations of Scientific Theories: The Vacuum Approach to the Problem of Motion in General Relativity', *Philosophy of Science*, **84**(5), pp. 1202–1214.

Linnemann, N., Read, J. and Teh, N. J. [2023]: 'The local validity of special relativity: the standard story articulated and defended', .

Matzner, R. A. [1968]: 'Almost Symmetric Spaces and Gravitational Radiation', *Journal of Mathematical Physics*, **9**(10), pp. 1657–1668.
<https://doi.org/10.1063/1.1664495>

Maudlin, T. [2019]: *Philosophy of Physics: Quantum Theory*, Princeton University Press.

Patton, L. and Curiel, E. [2023]: *Working Toward Solutions in Fluid Dynamics and Astrophysics: What the Equations Don't Say*, Springer Verlag.

Pitts, J. B. [2010]: 'Gauge-invariant localization of infinitely many gravitational energies from all possible auxiliary structures', *General relativity and gravitation*, **42**(3), pp. 601–622.

Pitts, J. B. [2022a]: 'General Relativity, Mental Causation, and Energy Conservation', *Erkenntnis*, **87**(4), pp. 1931–1973.

Pitts, J. B. [2022b]: 'On Two Slights to Noether's First Theorem: Mental Causation and General Relativity', in Y. Ben-Menahem (*ed.*), *Rethinking the Concept of Laws of Nature: Natural Order in the Light of Contemporary Science*, Springer.

Pitts, J. B. [2022c]: 'What represents space-time? And what follows for substantivalism vs. relationalism and gravitational energy?', in A. Vassallo (*ed.*), *The Foundations of Spacetime Physics*, Routledge.
<https://doi.org/10.4324

Pooley, O. [2017]: '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**, pp. 121101.
<https://link.aps.org/doi/10.1103/PhysRevLett.95.121101>

Read, J. [2016]: *Background independence in classical and quantum gravity*, University of Oxford (B.Phil. thesis).

Read, J. [2020]: ‘Functional Gravitational Energy’, *The British journal for the philosophy of science*, **71**(1), pp. 205–232.

Read, J. [2022]: ‘Geometric objects and perspectivalism’, in J. Read and N. Teh (*eds*), *The Physics and Philosophy of Noether’s Theorems*, Cambridge University Press.

Rueger, A. [1998]: ‘Local Theories of Causation and the a posteriori Identification of the Causal Relation’, *Erkenntnis*, **48**(1), pp. 25–38.

Szabados, L. B. [1992]: ‘On canonical pseudotensors, Sparling's form and Noether currents’, *Classical and Quantum Gravity*, **9**(11), pp. 2521–2541. <https://doi.org/10.1088/0264-9381/9/11/017>

Szabados, L. B. [2009]: ‘Quasi-Local Energy-Momentum and Angular Momentum in General Relativity’, *Living reviews in relativity*, **12**(1), pp. 4–4.

Szabados, L. B. and Tod, P. [2019]: ‘A positive Bondi-type mass in asymptotically de Sitter spacetimes’, *Classical and Quantum Gravity*, **32**(20), pp. ARTN 205011.

Trautman, A. [1962]: ‘Conservation laws in general relativity’, *Gravitation: An introduction to current research*, pp. 169–198.

Vassallo, A. [2019]: ‘Dependence relations in general relativity’, *European Journal for Philosophy of Science*, **10**(1).

Wald, R. M. [1984]: *General Relativity*, Chicago, USA: Chicago University Press.

Wallace, D. [2012]: *The emergent multiverse: quantum theory according to the Everett interpretation*, Oxford University Press.

Wallace, D. [2019]: ‘Who’s afraid of coordinate systems? An essay on representation of spacetime structure’, *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, **67**, pp. 125–136. <https://www.sciencedirect.com/science/article/pii/S1355219817300965>

Wallace, D. [2020]: ‘Fundamental and Emergent Geometry in Newtonian Physics’, *The British Journal for the Philosophy of Science*, **71**(1), pp. 1–32. <https://doi.org/10.1093/bjps/axx056>