D. RICKLES

WHO’S AFRAID OF BACKGROUND INDEPENDENCE?

ABSTRACT: Background independence is generally considered to be ‘the mark of distinction’ of general relativity. However, there is still confusion over exactly what background independence is and how, if at all, it serves to distinguish general relativity from other theories. There is also some confusion over the philosophical implications of background independence, stemming in part from the definitional problems. In this paper I attempt to make some headway on both issues. In each case I argue that a proper account of the observables of such theories goes a long way in clarifying matters. Further, I argue, against common claims to the contrary, that the fact that these observables are relational has no bearing on the debate between substantivalists and relationalists, though I do think it recommends a structuralist ontology, as I shall endeavour to explain.

1 INTRODUCTION

Everybody says they want background independence, and then when they see it they are scared to death by how strange it is ... Background independence is a big conceptual jump. You cannot get it for cheap... ([Rovelli, 2003], p. 1521)

In his ‘Who’s Afraid of Absolute Space?’ [1970], John Earman defended Newton’s postulation of absolute, substantival space at a time when it was very unfashionable to do so, relationalism being all the rage. Later, in his World Enough and Space-Time, he argued for a tertium quid, fitting neither substantivalism nor relationalism, substantivalism succumbing to the hole argument and relationalism offering “more promissory notes than completed theories” ([Earman, 1989], p. 195). More recently, in a pair of papers written with Gordon Belot [1999; 2001], substantivalism comes under attack again. This time the culprit is the background independence of general relativity, and the potential background independence of a future theory of quantum gravity. The claim is that were the successful future theory of quantum gravity shown to be background independent, then substantivalism would be rendered untenable for reasons of physics—thus providing a clear-cut example of Shimonyan ‘experimental metaphysics’ in action.\(^1\) Still more recently, in Volume 1 from this series [Dieks, 2006], Earman returns to his tertium quid idea, defending, again on the basis of (a manifestation of) background independence.

\(^1\) In fact, this is really just the hole problem again. In another paper, [Rickles, 2005b], I explicitly translated the hole argument into the framework of (background independent) loop quantum gravity, thus demonstrating that (this approach to) quantum gravity does not put the debate between substantivalists and relationalists on better ground than in the classical theory: substantivalists have nothing to fear from quantum gravity (not in the case of loop quantum gravity at any rate). I will aim to strengthen this conclusion further in this chapter.
independence, what I consider to be a \textit{structuralist} position which denies the fundamental existence of \textit{subjects} (in the sense of ‘bearers’ of properties), thus ruling out both relationalism and substantivalism [Earman, 2006a].

However, my primary target in this paper is the issue of what background independence is: only when this is resolved can we assess the claim that it might serve to settle debates over the ontology of spacetime (my secondary target). Let us begin by considering some basic metaphysical aspects of background structure, dependence, and independence, before firming the discussion up with the technical and definitional aspects. Ontological implications must wait until the final section.

2

\textbf{METAPHYSICS OF BACKGROUNDS}

Metaphysicians like to tell the following story to distinguish between physicalism and other non-physicalist positions:

\begin{quotation}
When God made the world did He lay out all the local physical matters of fact (properties at spacetime points) and the rest (causation, laws, modality, consciousness, etc.) followed, or did He then have to add these \textit{after} or \textit{in addition} to doing that?\footnote{Earman goes so far as to suggest an entirely new ontological category: a “\textit{coincidence occurrence}” ([Earman, 2006a], p. 16). This is close to what I take to be one of the main ontological implications of background independence; however, as I have already intimated, I couch matters in structuralist terms—see \S4. See also [Rickles, 2006] for a similar proposal drawn from the frozen formalism and problem of time in classical and quantum gravity, and [Rickles, Forthcoming] for a more general defense of the view on the basis of (gauge) symmetries in physics.}
\end{quotation}

Physicalists, of course, think that in fixing the local physical facts in a world He thereby fixed \textit{everything} there is to that world: all that there is is physical. \textit{Mutatis mutandis}, we can use this strategy to distinguish between positions on spacetime ontology too:

\begin{quotation}
When God made the world did He first create spacetime and \textit{then} add matter (particles, fields, strings, etc.) to it or did He create matter and \textit{thereby} fix the existence of spacetime?
\end{quotation}

Or, in other words, do spacetime, and spatiotemporal properties and relations, exist \textit{independently} of physical, material objects (particles, fields, strings, branes, etc...) or is the existence of some such objects \textit{necessary}\footnote{In other words, are causation, laws, modality, consciousness, etc., \textit{supervenient} on the local physical matters of fact (but not \textit{vice versa}), or do they constitute something ‘over and above’ these facts?} for their existence? Substantivalists will answer \textit{Yes} to the first disjunct and relationalists will answer \textit{Yes} to the second. Let us be clearer on exactly what is meant by these terms. Here I shall follow Sklar [1974] (himself followed by Earman and a generation of philosophers of physics) in taking substantivalism to be the position that views spacetime to be an entity which exists \textit{over and above} any material objects it might contain; or, in Earman’s words, “prior to the objects it contains” instead of being “nothing

\textit{Or just \textit{sufficient} if we wish to wage a war over ontological parsimony.}
but (might be constituted by, might be reducible to) the mutual relations among coexistent objects” ([Earman, 1989], p. 289).

This also captures much of the intuitive distinction between background independence and dependence: is spacetime (geometry) fixed ‘prior’ to the determination of the state of matter in the universe or does one need to know what the state of matter is ‘prior’ to the determination of spacetime geometry? Given this superficial similarity, the distinction between background independence and background dependence is often supposed to latch on to the distinction between relationalism and substantivalism: relationalism being committed to the former; substantivalism being committed to the latter. However, the ‘dual role’ of the metric field in general relativity rather muddies the waters here.

The schizophrenic nature of the metric field was viewed by Einstein as a necessary consequence of the equivalence principle, identifying inertia and weight: “the symmetric ‘fundamental tensor’ \( g_{\mu\nu} \) determines the metrical properties of space, the inertial behaviour of bodies in it, as well as gravitational effects” ([Einstein, 1918c], p. 241). Or, as Carlo Rovelli puts it: “What Einstein has discovered is that Newton had mistaken a physical field for a background entity. The two entities hypostatized by Newton, space and time, are just a particular local configuration of a physical entity — the gravitational field — very similar to the electric and the magnetic field” ([Rovelli, 2006], p. 27). In other words: “Newtonian space and time and the gravitational field are the same entity” (ibid.).

This duality, one expression of background independence in general relativity, has been responsible for much recent debate in the philosophy of spacetime physics. Again, it is seen to be implicated in the traditional debate between substantivalism and relationalism:

In Newtonian physics, if we take away the dynamical entities, what remains is space and time. In relativistic physics, if we take away the dynamical entities, nothing remains. As Whitehead put it, we cannot say that we can have spacetime without dynamical entities, anymore than saying that we can have the cat’s grin without the cat. ([Rovelli, 2006], p. 28)

The reason being that the gravitational field is dynamical and does the work of two in also supplying the structures that characterize spacetime. However, the proposed link to the substantivalism-relationalism debate is problematic. Rovelli is lumping all of the dynamical fields together, as being ontologically ‘on all fours’; but this is a mistake: we can remove all fields with the exception of the gravitational field and still have a dynamically possible world—i.e. there are vacuum

---

5 There are other, prima facie more substantive reasons for the alignment; however, these reasons ultimately fail as well: see [Rickles, 2005b; 2005a; 2006; Forthcoming] for the reasons why.

6 The still Machian Einstein of 1918 would not agree with this claim. He writes that “with [Mach’s Principle] according to the field equations of gravitation, there can be no G-field without matter” ([Einstein, 1918b], p. 34) — of course, this is where his \( \lambda \)-term appears, precisely in order to make the field equations compatible with Mach’s Principle. The field equations lose out, being transformed into \( G_{\mu\nu} - \lambda g_{\mu\nu} = -\kappa (T_{\mu\nu} - \frac{1}{2} g_{\mu\nu} T) \), which do not allow for empty (i.e. \( T_{\mu\nu} = 0 \)) spacetimes. The position I come to defend is not a million miles away from this: other fields are needed to form the gauge-invariant correlations (between field values) that provide the basic physical content of the theory.
solutions to Einstein’s field equations. But we cannot remove the gravitational field in the same way, leaving the other fields intact. This is an indication that there is something special about the gravitational field: it can’t be switched off; it is not just one field among many.

Hence, the substantivalist would be perfectly within her rights to claim ownership. But, so would the relationalist since there is this ambiguity over the ontological nature of the field\(^7\): spacetime or material object? I take this state of affairs (namely ‘joint ownership’ of the metric field) to lend substantial support to Robert Rynasewicz’s [1996] claim that the debate between substantivalism and relationalism is “outmoded” in this context. However, we are drifting somewhat from our brief, which is to get a grip on the concept of background independence (and its companions, background structure and background dependence).

Background structures are contrasted with dynamical ones, and a background independent theory only possesses the latter type—obviously, background dependent theories are those possessing the former type in addition to the latter type.\(^8\) Philosophers, and some physicists, will be more familiar with the term ‘absolute element’ in place of background structure, and the latter concept certainly soaks up a large part of the former. On the former concept, in his 1921 Princeton Lectures on the Theory of Relativity, Einstein writes:

> Just as it was consistent from the Newtonian standpoint to make both the statements, *tempus est absolutum, spatium est absolutum*, so from the standpoint of the special theory of relativity we must say, *continuum spatii et temporis est absolutum*. In this latter statement *absolutum* means not only “physically real,” but also “independent in its physical properties, having a physical effect, but not itself influenced by physical conditions. ([Einstein, 1921], p. 315)

There are three components here: a realist thesis, an independence thesis, and a non-dynamical thesis. Clearly the realist thesis can’t simply mean that space and time exist, for Leibniz too would surely assent to such a thesis in *some* sense. Instead, I take it to mean that space and time are *fundamental* in the sense that they do not supervene on any further, underlying objects, properties, or facts. The independence thesis just looks like a denial of relationism, while the non-dynamical thesis amounts to ‘absolute’ in something like the sense of Anderson’s notion of ‘absolute object’ ([Anderson, 1967], pp. 83-7)—this itself, of course, corresponds most closely to Newton’s notion of absolute in the sense of *immutability*, itself followed very closely by Einstein himself:

If Newton called the space of physics ‘absolute’, he was thinking of yet another property of that which we call ‘ether’. Each physical object influences and in general is influenced in turns by others. The

---

\(^7\) For example, the relationalist might, as Rovelli does, draw attention to the fact that “a strong burst of gravitational waves could come from the sky and knock down the rock of Gibraltar, precisely as a strong burst of electromagnetic radiation could” ([Rovelli, 1997], p. 193).

\(^8\) There is often a fair amount of slipping and sliding on this: there are *degrees* of background structure. Generally, one has in mind background *fields* rather than structures *per se*; that is, one is interested in the freedom (or not) from geometric-object fields on a manifold that are deemed ‘background’. Though the manifold itself appears as a background structure, this is generally not counted when assessing a theory’s background independence. This is a contentious point that we return to later—see, especially, footnote 10.
latter, however, is not true of the ether of Newtonian mechanics. The inertia-producing property of this ether, in accordance with classical mechanics, is precisely not to be influenced, either by the configuration of matter, or by anything else. For this reason, one may call it ‘absolute’. ([Einstein, 1999], p. 15)

Anderson prefers to call this “the principle of reciprocity”:

It is seen that the absolute elements of a theory effect the physical behaviour of a system. That is, a different assignment of values to the absolute elements would change the physical behaviour of the system. For instance, the assignment of different values to the metric might result in particle paths that are circles rather than straight lines. On the other hand, the physical behaviour of a system does not affect the absolute elements. An absolute element in a theory indicates a lack of reciprocity; it can influence the physical behaviour of the system but cannot, in turn, be influenced by this behaviour. This lack of reciprocity seems to be fundamentally unreasonable and unsatisfactory. We may express the converse in what might be called a general principle of reciprocity: Each element of a physical theory is influenced by every other element. In accordance with this principle, a satisfactory theory should have no absolute elements. ([Anderson, 1964], p. 192)

Lee Smolin too adopts a similar line, explicitly linking this notion of absolute with the notion of background. He writes that “[t]he background consists of presumed entities that do not change in time, but which are necessary for the definition of the kinematical quantities and dynamical laws” ([Smolin, 2006], p. 204). However, matters are not so simple as this. This rough ‘absolute elements’ way of defining background independence and background dependence is too vague to do any real work, and the various methods of firming things up face serious problems (as we shall see). Moreover, the peculiar nature of general relativity, replete with its treatment of the metric as a local dynamical variable, threatens to collapse the debate between substantivalism and relationalism. The interpretation of spacetime physics appears to be floundering.

Yet both debates, between background independence and background dependence and between substantivalism and relationalism, are believed by many physicists and a handful of philosophers to play a vital role in the search for a quantum theory of gravity. For example, in much of his recent work Lee Smolin (e.g. [2004; 2006]) defends the idea that background independence is a necessary piece of the quantum gravity puzzle: it is essential to solve the puzzles that quantum gravity raises that the geometry of spacetime is given as a solution of some equations of motion, rather than placed in the theory ‘by hand.’ But Smolin also argues that background independence uniquely supports relationalism, claiming that physicists “often take background independent and relational as synonymous” ([Smolin, 2006], p. 204). A big target in this paper is just this claim—a claim also made by Belot and Earman [1999; 2001] in order to prop up the listless body of the substantivalism/relationalism debate. Substantivalists needn’t be afraid of background independence any more than relationalists. However, ultimately both lose out to a structuralist position!
DEFINITIONS AND DISPUTATIONS

It is often claimed that the novelty of general relativity lies in its (manifest) ‘background independence.’ However, background independence is a slippery concept apparently meaning different things to different people. In this section we attempt to gain a firmer grip on this slippery customer by considering various elucidations of background independence that have been suggested. There is a clear core to the notion, and I argue that this core can be made clearer by connecting the concept of background independence to the nature of the observables in background independent theories.

Let us begin by presenting a general way of making sense of the various proposals—here I largely follow [Giulini, Forthcoming]. Let us specify a theory by writing down its laws as a set of equations of motion representing relations between the central objects of the theory. We get the following schema:

\[ \mathcal{E}[D, B] = 0 \]

Here \( D \) represents the dynamical structures (those that have to be solved to get their values, such as the electromagnetic field and the metric in general relativity—these represent the physical degrees of freedom of the theory, out of which the observables will be constructed) and \( B \) the background structures (those whose values are put in ‘by hand’, such as the topology and, in pre-general relativistic theories, the metric). Now let us represent the space of kinematically possible histories by \( K \). Then \( \mathcal{E}[D, B] = 0 \) selects a subset \( P \subset K \) of dynamically possible histories (or ‘physically’ possible worlds) relative to \( B \).\(^9\) Now, if there are no such \( B \)-s (or, rather, no \( B \)-fields) then the physically possible histories (the dynamics) is given by relations between the \( D \)-s (and, at least fiducially—i.e. in terms of the formal definition of the fields—the manifold, but the diffeomorphism symmetry washes this dependence away). This impacts on the observables of the theory, for the observables must then make no reference to the \( B \)-s, only to the \( D \)-s. This is the source of the claim that general relativity, and background independent theories, are relational: it simply means that the states and observables of the theory do not make reference to background structures.\(^10\)

\(^9\)As Wheeler puts it, “[k]inematics describes conceivable motions without asking whether they are allowed or forbidden. Dynamics analyses the difference between a physically reasonable and a disallowed history” ([Wheeler, 1964], p. 65).

\(^10\)Though, again, this does not include the manifold which is required for the (formal) definition of the dynamical fields. The inescapable presence of the manifold, in which dimension, topology, differential structure and signature are fixed independently of the equations of motion, leads Smolin to call general relativity only a “partly relational theory” ([2006], §7.4). However, the absence of background fields coupled with the symmetry of the manifold means that a displacement (via a diffeomorphism) of the dynamical fields with respect to it simply produces a gauge-equivalent representation of one and the same physical state. Elimination of these redundant possibilities (“surplus structure” in Redhead’s sense [1975]) further reduces the size of \( \mathcal{P} \), giving us the reduced space \( \overline{\mathcal{P}} = \mathcal{P}/\text{Diff}(\mathcal{M}) \). This ‘superspace’ contains points that are entire orbits of the gauge group, representing abstract ‘delocalized’ structures known as a “geometries”—see [Misner et al., 1973], p. 522. This is supposed to be a space...
This way of understanding a theory lets us recapitulate in a clearer way our earlier definitions of covariance and invariance. Let \( G \) be a group of spacetime symmetries that acts on \( K \) as \( G \times K \to K \)—i.e. elements of \( G \) map kinematically possible solutions onto kinematically possible solutions. We say that \( G \) is a symmetry group of the theory whose space of kinematically possible histories is \( K \) just in case \( P \) is left invariant by its action. Alternatively, and more usefully for what follows, we can express the distinction between covariance and invariance as—this is summarized schematically in fig. 1:

\[
\begin{align*}
\text{[COV]} \Rightarrow & \quad \mathcal{E}[D, B] = 0 \iff \mathcal{E}[g \cdot D, g \cdot B] = 0 \quad (\forall g \in G) \quad (2) \\
\text{[INV]} \Rightarrow & \quad \mathcal{E}[D, B] = 0 \iff \mathcal{E}[g \cdot D, B] = 0 \quad (\forall g \in G) \quad (3)
\end{align*}
\]

![Figure 1. How to understand covariance and invariance groups in a spacetime theory. Here, the fields take values in a vector space (or a more structured space). The diffeomorphisms drag fields along to new points. The equations of motion are of the form 'solve for \( D \) given \( B \).'](image)

fit only for relationalists; however, there are plenty of good arguments that show that the substantivalist has just as much right to occupy it—see [Pooley, Forthcoming] and [Rickles, Forthcoming] for more details.

\[\text{Cf.} \quad \text{Giulini [Forthcoming], p.6. I recommend that all philosophers of physics interested in background independence, and the difficulties in defining absolute objects, read this article: it is an exceptionally clear-headed review.}\]
As I said, in the context of general relativity $[\text{COV}] = [\text{INV}]$ since $\mathcal{B} = \emptyset$ (manifold aside). Of course, the fact that the manifold appears in the laws—and the absence of symmetry-reducing background fields (i.e. to reduce the effective symmetry group of the theory)—means that there will be surplus structure: the localization of the manifold of the dynamical fields is pure gauge.\footnote{This difference corresponds, then, to that between ‘passive’ and ‘active’ diffeomorphism invariance. As Rovelli puts it: “A field theory is formulated in manner invariant under passive diffs (or change of co-ordinates), if we can change the coordinates of the manifold, re-express all the geometric quantities (dynamical and non-dynamical) in the new co-ordinates, and the form of the equations of motion does not change. A theory is invariant under active diffs, when a smooth displacement of the dynamical fields (the dynamical fields alone) over the manifold, sends solutions of the equations of motion into solutions of the equations of motion” (Rovelli, 2001, p. 122). We will call the former general covariance and the latter diffeomorphism invariance—Earman (2006b) calls the latter substantive general covariance, on the understanding that it amounts to a gauge symmetry, as we have assumed.}

All this symmetry affects the dynamics so that a standard Hamiltonian or Lagrangian formulation is not possible. Respectively, the canonical variables are not all independent (being required to satisfy identities known as constraints: $\phi(q, p) = 0$) and the Euler-Lagrange equations are not all independent. These identities serve to ‘constrain’ the set of phase space points that represent genuine physical possibilities: only those points satisfying the constraints do so, and these form a subset in the full phase space known as the ‘constraint surface’.

As I also said, this has an impact on the form of the observables—and this is terribly important for the quantization of the theory. Since a pair of dynamical variables (not observables) that differ by a gauge transformation are indistinguishable, corresponding to one and the same physical state of affairs, the observables ought to register this fact too: that is, the observables of a gauge theory should be insensitive to differences amounting to a gauge transformation—as should the states in any quantization of such a theory: i.e. if $x \sim y$ then $\Psi(x) = \Psi(y)$.\footnote{It seems that Einstein was aware of this implication soon after completing his theory of general relativity, for he writes that “the connection between quantities in equations and measurable quantities is far more indirect than in the customary theories of old” ([Einstein, 1918a], p. 71).} Where ‘$O$’ is a dynamical variable, ‘$O$’ is the set of (genuine) observables, $x, y \in \mathcal{P}$, and ‘$\sim$’ denotes gauge equivalence, we can express this as:

\begin{equation}
O \in O \iff (x \sim y) \supset (O(x) = O(y))
\end{equation}

Or, equivalently, we can say that the genuine observables are those dynamical variables that are constant on gauge orbits ‘$[x]$’ (where $[x] = \{x : x \sim y\}$):

\begin{equation}
\forall [x], O \in O \iff O[x] = \text{const}.
\end{equation}

Most of the work done on finding the observables of general relativity is done using the $3 + 1$ projection of the spacetime Einstein equations. That is, the constraints are understood as conditions laid down on the initial data\footnote{Note that John Wheeler refers to constraints as “initial value equations” ([Wheeler, 1964], p. 83). This terminology gets one closer to the physical meaning of the constraints.} ($\Sigma, h, K$) when we project the spacetime solution onto a spacelike hypersurface $\Sigma$—here, $h$ is a Riemannian
metric on \( \Sigma \) and \( K \) is the extrinsic curvature on \( \Sigma \). I won’t go into the nitty gritty details here, but it turns out that the Hamiltonian of general relativity is a sum of constraints on this initial data (of the kind that generate gauge motions, namely 1st class)—hence, the dynamics is entirely generated by constraints and is therefore pure gauge.\(^{15}\)

This formulation allows us to connect the characterization of the observables up to the dynamics (generated by constraints \( \mathcal{H}_i \)) more explicitly:

\[
(6) \quad \mathcal{O} \in \mathcal{O} \iff \{ \mathcal{O}, \mathcal{H}_i \} \approx 0 \quad \forall i
\]

In other words, the observables of the theory are those functions that have weakly vanishing (i.e. on the constraint surface) Poisson brackets with all of the first-class constraints. These are the gauge-invariant quantities. A pressing problem in general relativity—especially pressing for quantum gravity—is to find suitable entities that satisfy this definition. There are at least two types that fit the bill: highly non-local quantities defined over the whole spacetime\(^{16}\) and (differently) non-local, ‘relational’ quantities built out of correlations between field values. There seems to be some consensus forming, at least amongst ‘canonical relativists’, that the latter type are the most natural.

I return to the interpretation of these correlational observables in §4. Let us now consider a number of standard takes on the question of what background independence amounts to. We assess two main ways of doing this—utilizing general covariance and diffeomorphism invariance—before considering J. L. Anderson’s proposal for firming up the latter approach and discussing the connection to observables.

As we saw above, general covariance simply refers to the fact that when we hit a solution with an arbitrary diffeomorphism, we get another solution back. That is to say, the equations of motion are covariant with respect to diffeomorphisms. This amounts to a a carrying along of all of the fields. Covariance is not so restrictive as invariance (or “symmetry” as Anderson calls it). The former just says that if \( M \) is a solution then so is \( M' = g(M) \) and \( g \) is an element of the covariance group, the group that preserves the form of the laws (the equations of motion). The theory is said to be \( g \)-covariant. But this is just a constraint on the form of the theory, not on its physical content. In other words, general covariance in this sense is simply a property of the formulation of the theory. This is, of course, just what Kretschmann taught Einstein soon after general relativity was written down in its final form. The problem is that (general) covariance just means that the equation

\(^{15}\)This turns out to be behind the two worst conceptual problems of general relativity: the hole argument and the frozen formalism problem. For details on these connections see [Belot and Earman, 1999; 2001; Rickles, 2006]. Earman [2003] gives a splendid presentation of the relationship between the constrained Hamiltonian formalism and gauge, including their implications for time and change.

\(^{16}\)There is a proof (for the case of closed vacuum solutions of general relativity) that there can be no local observables at all [Torre, 1993]—‘local’ here means that the observable is constructed as a spatial integral of local functions of the initial data and their derivatives.
of motion is *well-defined* in the sense of differential geometry: the equation need simply ‘live’ on the manifold.

Recall how coordinatization occurs. Firstly, we associate points of the manifold with $\mathbb{R}^4$ so that each $x \in \mathcal{M}$ gets four numbers $\{x^\mu\} (\mu = 1, 2, 3, 4)$ associated to it. We can do this in many ways, as mentioned above. We might use the assignment $\{x'^\mu\} \rightarrow y \in \mathcal{M}$ instead. Because these numbers are assigned to the same point there will be some relationship between the coordinate systems:

$$\begin{align*}
x'^\mu &= x'^\mu(a')
\end{align*}$$

Given the differential structure of the manifold, we get an infinitely continuously differentiable function between the thus related coordinate systems (with a similarly differentiable inverse).\(^{17}\) This is a diffeomorphism passively construed; it is gauge in a very trivial sense, as Wheeler says: “How one draws coordinate surfaces through space-time is a matter of paperwork and bookkeeping, and has nothing to do with the real physics” ([Wheeler, 1964], p. 81). This goes for any reasonable spacetime theory. Hence, general covariance understood in these terms does not have the power to distinguish between spacetime theories, and if background independence is supposed to distinguish general relativity from previous theories, then general covariance cannot underwrite it.

Hence, *any* spacetime theory written in terms of geometric objects on the manifold will be generally covariant in the sense of having $\text{Diff}(\mathcal{M})$ as its covariance group. However, *invariance* is a much stronger requirement that picks out a subgroup of the covariance group; this says that if $M$ is a solution then so is $M'$ where $M' = g(M)$ and $g$ and now $g$ is an element of a subgroup of the covariance group that preserves the absolute elements. The theory is said to be $g$-invariant. Hence, in one we map all of the objects, in the other we only map the dynamical objects, whilst preserving the background structure.\(^ {18}\)

This is the standard view: rather than considering the background fields to be transformed along with the dynamical fields, we view the diffeomorphisms in an active way, as shifting the dynamical fields relative to the same background fields. Thus, Lee Smolin writes that “[g]eneral coordinate invariance [general covariance—DR] is not the same thing as diffeomorphism invariance, and it is the latter, and not the former, that is the key to the physical interpretation of the theory”. He goes on to say that

\(^{17}\)It is useful to think of the pair of coordinate systems as being like a pair of languages and as the particular coordinates assigned to some particular point as being like nouns in the language referring to a particular object. A translation manual between the languages would be analogous to the differentiable functions relating distinct coordinate descriptions of one and the same point; in this case we would have distinct words referring to the same object, these words being inter-translatable.

\(^{18}\)Generally, because the manifolds in general relativity are coordinatized by gluing patches together, the function will be evaluated on the overlap between coordinate systems.

\(^{19}\)The story then goes: special relativity cannot be diffeomorphism invariant—i.e. it cannot have $\text{Diff}(\mathcal{M})$ as its symmetry group—because the imposition of the Minkowski metric reduces the invariance group to a subgroup of the covariance group, namely the Poincaré group of isometries of this metric, of Minkowski spacetime. This smaller group is the largest that preserves the (background) structure of Minkowski spacetime; there are clearly elements of $\text{Diff}(\mathcal{M})$ that would not do so.
with the introduction of explicit background fields any field theory can be written in a way that is generally coordinate invariant. This is not true of diffeomorphism invariance, which relies on the fact that in general relativity there are no non-dynamical background fields. Diffeomorphisms, in contrast to general coordinate transformations, are active transformations that take points to other points, so that diffeomorphism invariance is, explicitly, the statement that the points are not meaningful. Both philosophically and mathematically, it is diffeomorphism invariance that distinguishes general relativity from other field theories. (Smolin, 2003), p. 234).

There are at least two ways in which this misses the mark. Firstly, one can retain the physical content of diffeomorphism invariance without disposing of points: either one can adopt a Kretschmann-Bergmann-Komar ‘intrinsic coordinates’ method ([Kretschmann, 1917; Bergmann, 1977; 1961; Bergmann and Komar, 1972]—see also §4), or else one can view diffeomorphism invariance as imposing a constraint on the form of the observables of a theory, so that it is true that “points are not meaningful” only in the sense that from the point of view of the physics, as encoded in the observable content of the theory, there is an ‘indifference’ to the points of the manifold (i.e. as to which point plays which role). Both are compatible with their being points. Secondly, and more problematic for our purposes, is that this faces a Kretschmann-type objection too: any background field can be made dynamical by making it satisfy some equations of motion, however physically vacuous they might happen to be. Hence, unless we have some other way of making sense of the distinction between background and dynamical fields, then this account fails in the same way the general covariance account fails—fortunately, Anderson provides just such a method, but first let us go through the details of why this account fails as it stands.

The objection is that we are free to extend the invariance group to the covariance group by making any background fields into dynamical ones, thus collapsing the distinction between invariance and covariance groups. If there are no background fields then the invariance group automatically becomes identical to the covariance group (i.e. the diffeomorphism group). In the case of a specially relativistic theory, say, in order to preserve the structure of Minkowski spacetime we would have to impose a condition of flatness on the metric. But, of course, this makes the metric dynamical (in the sense of satisfying equations of motion)! The problem is, this all depends upon the availability of some way of distinguishing between absolute and dynamical fields, and so far we simply have an intuitive notion. Clearly if this intuitive notion amounts to ‘being solved for’ then we can make special relativity background independent, which then conflicts with our basic intuitions about what background independence is.

Take the following stock example of a massless scalar field on Minkowski spacetime:

\[
\Box \phi \equiv \eta^{\mu\nu} \nabla_\mu \nabla_\nu \phi = 0
\]

All we do here is replace the background metric with a general metric and make the new metric obey a ‘flatness condition’.
\(\eta_{\mu\nu} \rightarrow g_{\mu\nu}\)  \hspace{1cm} (9)
\(\Box \phi \equiv g^{\mu\nu} \nabla_\mu \nabla_\nu \phi = 0\)  \hspace{1cm} (10)
\(\text{Riem}[g] = 0\)  \hspace{1cm} (11)

Like the generally relativistic case, we now appear to have no background fields! If diffeomorphism invariance is what underwrites background independence then the latter cannot be what makes general relativity special.\(^{20}\) Hence, it appears that we have made a specially relativistic theory generally relativistic!

Anderson ([1964], pp. 182-3) complains about this procedure on the grounds that the way the general metric was introduced was physically unmotivated: there is no need to have a general metric since nonflat metrics are not considered. Presumably this is similar to what Einstein had in mind when he complained that although it was possible to reformulate other spacetime theories in a generally covariant way, this does not produce their simplest formulation: “Among two theoretical systems, both compatible with experience, one will have to prefer the one that is simpler and more transparent from the point of view of the absolute differential calculus” ([Einstein, 1918b], p. 34). There is a lot of room for the metric to move with a general metric that is not being occupied in these reformulations; hence, no additional content is being added by using it. Anderson argues that the expansion of the invariance group of a theory (to encompass a larger covariance group of which it is usually a subgroup) reveals pre-existing absolute objects in the theory. The metric in special relativity is an absolute object whose “existence was masked by the fact that [it] had been assigned [a] particular [value]” ([Anderson, 1964], p. 192).

Anderson provides the split between background and dynamical fields so that the diffeomorphism invariance definition of background independence can do its work. What Anderson proposes is that absolute elements (called “absolute objects” in [Anderson, 1967]), understood as variables whose “determination is entirely independent of other physical objects of the theory” ([1964], p. 186), serve to define the relativity principle associated with some theory—that is, the principle stating that the theory’s laws are invariant under the invariance group. Let us work through Anderson’s proposal to see how this works and how it contributes to the problem of defining background independence.

He begins, as we did above, by specifying a theory as a set of functional relations (i.e. equations of motion) between the independent variables of the theory (particles, fields, fluids, strings, branes, etc.):

\[\mathcal{L}_i(y_A) = 0\]

\(^{20}\)As Giulini notes ([Giulini, Forthcoming], pp. 13-4), there are in fact problems with this example: if we consider the (reasonable) requirement that our equations of motion have to be the Euler-Lagrange equations for some action principle then we find that the action principle delivering Eqs. 10 and 11 generates a bigger solution space than that of Eq. 8. The two are not equivalent formulations of the same theory.
Anderson seeks to find a way of identifying the background structures in a theory that is so specified. The idea is to inspect all of the invariant functions (the *genuine* observables) under the theory’s covariance group that can be constructed from various subsets of the variables \( y_A \), and to then see if the values of these functions are uniquely determined by Eq.\,12 alone (independently of additional conditions). Those values of the functions that are determined independently of the values of others are deemed absolute.

There is something *prima facie* rather peculiar about Anderson’s analysis in that it implies that the metric \( g_{\mu\nu} \) of general relativity constitutes an absolute element (with the Lorentz group as its invariance group) in the vacuum case, because it is uniquely determined (up to diffeomorphism) *independently of other physical objects*, but not in the matter-present case, where it is determined by other physical objects in the theory. Hence, it isn’t an absolute matter whether the metric in general relativity is an absolute object or not; rather, it depends on whether there are other fields present, and so on which field equations are appropriate. This is at odds with what we might expect, namely a definition of background independence that renders general relativity background independent *simpliciter*. However, given what I had to say about the observables—i.e., that they are correlations between fields values—I think this is what we *should* say.\(^\text{21}\)

Background independence is, then, defined using this machinery: a theory is background independent just in case it contains no absolute elements. This lines up with the diffeomorphism invariance account, for a diffeomorphism invariant theory will have no background structures; this is how we get the identity between the covariance and invariance groups. This clearly renders general relativity background independent; its covariance group is indeed identical to its invariance group (or its ‘relativity group’). The method gets the relativity principles for other space-time theories right too.

If we turn what were originally background fields into dynamical fields, by making them obey equations of motion (in the sense of Anderson), then they will enter into the definition of the observables, since we are understanding the \( \mathcal{D} \) to be the ‘ingredients’ of the observables. This is how we end up with hole-type problems: background fields—however they are tweaked in an attempt to make them dynamical—introduce unobservable (‘unphysical surplus’) content into the physical structure (as given by the \( \mathcal{D}s \)). This can be seen explicitly if we consider how the born-again dynamical fields look in the solution space of the theory. \( \mathcal{P} \) is our solution space, and since we are taking the theory to be diffeomorphism invariant it will carry an action of the gauge group \( \text{Diff}(\mathcal{M}) \). The dynamical fields serve to separate the points of \( \mathcal{P} \). However, there will be a redundancy in the labeling since the diffeomorphism invariance allows us to construct solutions from solutions by acting on them with elements of \( \text{Diff}(\mathcal{M}) \). For each solution of the original equations we now have an orbit of solutions. If we understand the diffeomorphism

\(^{21}\)Compare this with the Einstein quote I give on p.\,17. I think this clearly shows that Einstein would have sided with Anderson on this point.
symmetry as a gauge freedom then this will be a gauge orbit. This gives us a potential way to further detrivialize this approach, for we can see that the flatness condition forces the value of $g_{\mu \nu}$ to be the same in each and every orbit. Hence, if we identify the gauge orbits then we have just one state here—I am shelving complications to do with locality; for more details see Giulini ([Forthcoming], §2.5).

This is what it means to say the metric is an absolute object: something that is the same in every solution. But the inability to distinguish between orbits is the definition of an unobservable here too, so we have the connection between background independence and observables that we were looking for (and the connection between background structures and non-observability). Moreover, it matches what we intuitively mean by background independence.

4
IMPLICATIONS FOR THE ONTOLOGY OF SPACETIME

Diffeomorphism invariance makes local observables an impossibility. Since there clearly are local degrees of freedom, and these are what we observe, we need some notion of local observable that does not make reference to spacetime geometry. That is, we need a background independent notion of local observable. The obvious (and indeed, only thing to do) is to use physical degrees of freedom to localize. The observables so localized are relational.

Calling those dynamical variables whose motion can be uniquely determined by the field equations the “true observables”, Anderson writes that:

A unique state of the system is ... specified by giving, at some instant of time, values of the true observables and their first time derivatives. In a sense, these true observables are the physical meaningful “coordinates” of the system. ([Anderson, 1958], p. 1197)

These true observables are the gauge-invariant quantities I mentioned earlier. Earman asks: “Does the gauge-invariant content of GTR characterize a reality that

---

22 That is, the value is not just constant on gauge orbits—which is part and parcel of being a good observable—it is constant across gauge orbits too. No observable can distinguish between such orbits; hence the structure is unobservable.

23 There are problem cases that remain, as discussed, for example, in [Pitts, 2006]—namely, the so-called ‘Jones-Geroch counterexample’ which apparently shows that the 4-velocity of a ‘cosmic dust field’ counts as a background structure (according to the Anderson-Friedman analysis). The problem stems from Michael Friedman’s’ modification of Anderson’s identification of background structures as those with single $\text{Diff}(\mathcal{M})$ orbits. Friedman argues that the condition should be made local in order to get at the notion that background structures do not correspond to local degrees of freedom. To achieve this he counts as background structures fields that are locally diffeomorphism equivalent—this condition is satisfied when there is a diffeomorphism mapping neighbourhoods (of any any manifold point) to neighbourhoods, such that two fields restricted to the neighbourhoods (connected by a carry-along) take on the same values. The problem is that any pair of nowhere vanishing vector fields will always satisfy this condition and, therefore, always count as background structures. The absurd conclusion is that any diffeomorphism invariant vector field theory will automatically be branded background dependent. Utilizing the observables can help here, at least in the present case: the observables will register the physical fact that such fields will generally not cover the whole of spacetime.
answers the relationist’s dreams, or do the terms of the absolute-relational controversy no longer suffice to adequately describe what Einstein wrought?” (Earman, 2006a, p. 10). Earman answers Yes to the latter disjunct, and No to first. He proposes to put an entirely new ontological scheme, based on ‘coincidence occurrences’, in place of the absolute and relational positions. As Earman points out, “a coincidence occurrence consists in the corealization of values of pairs of (non-gauge invariant) dynamical quantities” (Earman, 2006a, p. 16). Earman thinks that this new conception of physical quantities signals the necessity of a shift from the traditional ‘subject-predicate’-based ontologies, such as substantivalism and relationalism. As I said earlier, I think this is the right thing to say; however, I would spell it out rather differently, in terms of structuralism. Rovelli’s framework of partial and complete observables provides the formal underpinning.

Firstly, how might relationalism and substantivalism get a foothold in this background independent context? According to the relationalist (about motion) all motion is relative motion. But motion relative to what? The gravitational field? But if it is the gravitational field, then we face a problem in GR (and background independent theories in general): is this field spacetime or matter? Einstein, and Rovelli, claim that the gravitational field should be identified with spacetime. Here we see that both positions can get a foothold on the ontological rock face of general relativity; the substantivalist can lay claim to the same object against which relative motion occurs. The same goes for localization, which is, I suspect, more what Earman has in mind: if localization is relative to the gravitational field, then both substantivalists and relationalists (in the ontological sense) can get a foothold.

Matters have clearly degenerated (pardon the pun) to the point where this division is no longer doing any real work.

But we can say more. I mentioned above that reduction (i.e. the elimination of symmetry) was supposed to be implicated. The idea here is that the natural representational tool for relational spacetime is the geometry rather than individual metrics on the manifold:

[T]he basic postulate that makes GR a relational theory is [that] ... [a] physical spacetime is defined to correspond, not to a single \((M, g_{ab}, f)\), but to an equivalence class of manifolds, metrics, and fields under the action of \(\text{Diff}(M)\) (2006, p. 206).

The idea here, then, is that removing the symmetries (by ‘modding out’ by the diffeomorphisms) is taken to correspond to relationalism; or, in other words, that relationalism is reductionism. This is tantamount to the gauge-invariance view. It poses no obstacle for the substantivalist; there are a variety of ways to accept it, most of which amount to a denial of haecceitism of some sort or another (i.e. the claim that there can be worlds that differ non-qualitatively)—see Pooley [Poo-ley, Forthcoming] for more details. There is no necessity gluing haecceitism and substantivalism together, and a relationalist can just as well be an haecceitist.

Like Rovelli and Smolin, Wheeler dismisses the points of space as “[m]ere baggage”. The coordinate representations we use hide the real, objective reality: the geometries. Hence, a geometry is an abstract object that encodes the intrinsic
features of a space: it stands one-to-many with localized metrics. Again, there
is no reason why the substantivalist shouldn’t say that this intrinsic structure is
what they mean by spacetime and where their ontological commitments lie. On
the other side of the coin it is possible that the substantivalist can retain points in
the face of the gauge freedom. As Robert Dicke remarks, speaking on behalf of J.
L. Synge:

general relativity describes an absolute space … certain things are measurable about this space in an
absolute way. There exist curvature invariants that characterize this space, and one can, in principle,
measure these invariants. Bergmann has pointed out that the mapping of these invariants throughout
space is, in a sense, labeling of the points of this space with invariant labels (independent of coordinate
system). These are concepts of an absolute space, and we have here a return to the old notions of an
absolute space. ([Dicke, 1964], p. 124-5)

Here the idea is to get a set of coordinate conditions that allow one to define a set
of intrinsic coordinates. One constructs the complete set of scalars from the metric
and its first and second derivatives, which for the matter-free case leaves four non-
zero scalars that take different values at different points of the manifold. Hence,
one achieves a complete labeling of the manifold in an intrinsic gauge-invariant—
this follows from the fact that we are dealing with scalars which do not change
their values under diffeomorphisms. These points can then be used to localize
quantities which become gauge-invariant as a result of the gauge-invariance of the
scalars. For Synge, the only difference between this space and Newton’s is that
the geometric properties of the Einsteinian space are “influenced by the matter
contained therein”—that is, the latter is background independent. Of course, since
we are dealing with invariants of the metric here, it is open to the relationalist to
call this a material field. So continues the interminable tug-of-war!

I think this is evidence in favour of the view that the time has come to forget
about the ‘debate’ between substantivalism and relationalism, and focus on an al-
ternative. Here I argue that structuralism offers a suitable alternative. The position
involves the idea that physical systems (which I take to be characterized by the
values for their observables) are exhausted by extrinsic or relational properties:
they have no intrinsic properties at all! This is a consequence of background in-
dependence coupled with gauge invariance. This leads to a rather odd picture in
which objects and structure are deeply entangled in the sense that, inasmuch as
there are objects, any properties they possess are structurally conferred: they have
no reality outside the correlation. What this means is that the objects don’t ground
the structure; they are nothing independently of the structure, which takes the form
of a (gauge-invariant) correlation between (gauge variant) field values. We can
sum this up by paraphrasing one of Hermann Minkowski’s infamous remarks:

24There is kinship here with Eddington who writes that “the significance of a part cannot be dis-
associated from the system of analysis to which it belongs. As a structural concept the part is a symbol
having no properties except as a constituent of the group-structure of a set of parts” ([Eddington, 1958],
p. 145); and later, “a structure does not necessarily imply an X of which it is the structure” ([Eddington,
1958], p. 151).
Henceforth spacetime by itself and matter by itself are doomed to fade away into mere shadows, and only some kind of union between the two can preserve their independent reality!

This admittedly rather wild-sounding metaphysics can be made more precise through the use of Rovelli’s framework of partial and complete observables.

A partial observable is a physical quantity to which we can associate a measurement leading to a number and a complete observable is defined as a quantity whose value (or probability distribution) can be predicted by the relevant theory. Partial observables are taken to coordinatize an extended configuration space $Q$ and complete observables coordinatize an associated reduced phase space $\Gamma_{red}$. The “predictive content” of some dynamical theory is then given by the kernel of the map $f : Q \times \Gamma_{red} \rightarrow \mathbb{R}^n$. This space gives the kinematics of a theory and the dynamics is given by the constraints, $\phi(q^a, p_a) = 0$, on the associated extended phase space $T^*Q$. The content appears to be this: there are quantities that can be measured whose values are not predicted by the theory. Yet the theory is deterministic because it does predict correlations between partial observables. The dynamics is then spelt out in terms of relations between partial observables. Hence, the theory formulated in this way describes relative evolution of (gauge variant) variables as functions of each other. No variable is privileged as the independent one (cf. [Montesinos et al., 1999], p. 5). The dynamics concerns the relations between elements of the space of partial observables, and though the individual elements do not have a well defined evolution, relations between them (i.e. correlations) do: they are independent of coordinate space and time.

The interpretation here is as follows: $\phi = T$ is a partial observable parametrizing the ticks of a clock (laid out across a gauge orbit), and $f = \alpha$ is another partial observable (also stretching out over a gauge orbit). Both are gauge variant quantities. A gauge invariant quantity, a complete observable, can be constructed from these partial observables as:

(13) $O_{f,T}(\tau, x) = f(x')$

These quantities encode correlations. They tell us what the value of a gauge variant function $f$ is when, under the gauge flow generated by the constraint, the gauge variant function $T$ takes on the value $\tau$. This correlation is gauge invariant. These are the kinds of quantity that a background independent gauge theory like general relativity is all about. We don’t talk about the value of the gravitational field at a point of the manifold, but where some other physical quantity (say, a value of the electromagnetic field) takes on a certain value. Once again, we find that Einstein was surprisingly modern-sounding on this point, writing that “the gravitational field at a certain location represents nothing ‘physically real,’ but the gravitational field together with other data does” ([Einstein, 1918a], p. 71).

Now here I would agree with Einstein and disagree with Rovelli about the interpretation of these correlations. Rovelli claims that “the extended configuration space has a direct physical interpretation, as the space of the partial observables”
Both spaces—the space of genuine (complete) observables and partial observables—are invested with physicality by Rovelli; the partial observables, in particular, are taken to be physical variables. Einstein argues that only the correlation is physically real. In this he is clearly followed by Stachel [1993] who argues that the kinematical state space of a background independent theory like general relativity has no physical meaning prior to a solution (so that only the dynamical state space is invested with the power to represent genuine physical possibilities; kinematics then being in this sense derivative).

It is for this reason that I think structuralism can help with the interpretation of background independent, gauge-invariant theories—that is, we don’t need to go as far as Earman in postulating a whole new ontological category. Recall that epistemic structural realism argues that the best we can hope for is to get to know structural aspects of the world, since we only ever get to observe relational properties rather than intrinsic ones (in our experiments and so on). However, in a background independent gauge theory like general relativity we have seen that the physical observables just are relational quantities: this is all there is! In other words, there’s nothing ‘underneath’ the relational properties (as encoded in the $D$-fields), so that these exhaust what there is, leading to an ontological structuralism. This is why we face the problems regarding the ‘subject-predicate’-style ontologies that Earman mentions: there are no independent subjects that are the ‘bearers’ of properties and the ‘enterers’ of relations. Hence, unless one can have objects without intrinsic properties (and I don’t think this is a metaphysically healthy route to follow), we should follow Earman’s lead, and I say that this journey will lead us to some variant of ontic structural realism.

5 CONCLUSION

I have argued that we can make good sense of background structure and background independence by following an Anderson-style account (involving the view that background structures have single $\text{Diff}(\mathcal{M})$ orbits) and utilizing the appropriate gauge-theoretic definition of ‘observable’. These set up a connection between Anderson’s idea and the intuitive notion that background structures are not the kinds of thing we can measure, and not the kinds of things that can ground things (fields values and so on) we might wish to measure. The ontological implications of background independence, so conceived, are not what is often claimed:

$25$Hence, we have here an empirical argument for ontic structural realism that evades the standard ‘no relations without relata’ objection. The relations are the correlations here (the gauge invariant, complete observables), and the ‘relata’ would be the gauge variant, partial observables. But the partial observables being gauge variant do not correspond to physical reality (at least not in any fundamental sense): only the complete observables do. We cannot decompose the correlations in an ontological sense, though we clearly can in an epistemic sense—indeed, the correlates constitute our ‘access points’ to the more fundamental correlations.
relationalism is not uniquely supported. Substantivalists too can uphold their interpretation in the context of background independent theories. However, aspects of the observable content of background independent theories was shown to cause problems for both relationalism and substantivalism. I argued that these aspects recommend a structuralist position.

BIBLIOGRAPHY


