What is the value in an intrinsic formalism?

Eleanor March*

Abstract

I discuss the distinction between extrinsic and intrinsic approaches to reformulating a theory with symmetries, and offer an account of the special value of intrinsic formalisms, drawing on a distinction between which mathematical expressions are meaningful within an extrinsic formalism and which are not.

1 Introduction

Caspar Jacobs (2022) has recently considered afresh the motivation vs. interpretation debate about symmetries (originally due to Møller-Nielsen (2017)). Very roughly, the interpretationalist says that it is legitimate *ab initio* to interpret symmetry-related models (SRMs) of a theory as representing the same physical state of affairs, even in the absence of a metaphysically perspicuous characterisation of their common ontology, whereas the motivationalist denies this.[1] Jacobs takes up the question of what it means for a characterisation of the common ontology of SRMs to be 'metaphysically perspicuous.' He argues that this is captured by the demand for an intrinsic formalism, in the sense of Field (2016).[2] By contrast, extrinsic formalisms—in which the invariant content of SRMs is captured using equivalence classes of symmetry-variant quantities,[3] for example, by quotienting the theory's space of models under the symmetry group in question[4]—do not provide a similarly metaphysically perspicuous characterisation of a theory's ontological commitments. Instead, they only provide an effective decision procedure for determining whether the theory is committed to some piece of structure or other.

Or this is the story Jacobs tells. However, whilst I am in agreement with Jacobs that intrinsic formalisms are more metaphysically perspicuous than extrinsic formalisms, I have two worries about his argument. The first is that Jacobs does not clearly distinguish between external sophistication—an approach to capturing the invariant content of SRMs in which one stipulates that SRMs are to 'count' as isomorphic—and the provision of an extrinsic formalism, which, as I will explain, should really be thought of as an instance not of external sophistication but reduction or internal sophistication. As we will see, this means that many of Jacobs' points which are supposed to count against extrinsic formalisms, and therefore in favour of intrinsic formalisms, fail to latch onto the intrinsic-extrinsic distinction, since they are really criticisms of external sophistication.

The second worry is that Jacobs' positive argument that only intrinsic formalisms are metaphysically perspicuous is tied very closely to a view on which interpreting a theory is primarily a matter of identifying its ontology. On this way of thinking, we get a handle on what a theory says the world is like, first and foremost, by getting a handle on what things are in the world,

*Faculty of Philosophy, University of Oxford. eleanor.march@philosophy.ox.ac.uk

[1] Though in light of recent work by Luc (2023), these positions are probably best understood as two extremes of a more-nuanced spectrum of motivationalist vs. interpretationalist views.

[2] Though note that some motivationalists may disagree here, i.e. on whether an intrinsic formalism is necessary to give a metaphysically perspicuous characterisation of the common ontology of SRMs. In particular, I would expect this to be the case for e.g. motivationalists who are fans of the Kleinian approach to geometry.

[3] Following Jacobs (2021b) and Wallace (2019), I take using equivalence classes to subsume coordinate-based approaches.

[4] Recall that given a space X on which a group G acts (from the left), one defines the quotient space X/G as the space of orbits of G, i.e. the space of equivalence classes $[x]$, $x \in [x]$ iff $gx \in [x]$ for all $g \in G$, which (providing the group action of G on X is suitably well-behaved, e.g. perhaps it is faithful) inherits the same kind of structure as the original space (i.e. if X is a manifold, or Hilbert space, or whatever, then so is X/G).

according to that theory. Thus, Jacobs says, intrinsic formalisms are valuable because they allow one to 'read off' this ontology from the theory's formalism.[5]

This is fair enough. But what I want to point out here is that this conception of theory interpretation is not compulsory. There are many other ways of getting a handle on what a theory says the world is like which do not proceed primarily by identifying that theory's ontology, and if one thinks that one of these is the way to go, then the idea that one can 'read off' the ontology of intrinsic formalisms is by-the-by, as far as metaphysical perspicuity is concerned. For example, one way to get a handle on what a theory says the world is like is by cataloguing what kind of things the theory is able to say about the world. On this kind of view, one gets a handle on what e.g. Maxwellian spacetime $\langle M, t_a, h^{ab}, \circlearrowleft\rangle$ is like not by saying that it is ontologically committed to $2^{\aleph_0}$ spacetime points, two degenerate metrics, and a standard of rotation, but by saying: it is the kind of spacetime structure in which fluids have a well-defined state of rotation, and infinitesimally-separated fluid elements a well-defined relative acceleration, but no absolute state of acceleration. Or, another way to get a handle on what a theory says the world is like is by cataloguing what kind of inferences it licences. For example, in general relativity or Newton-Cartan theory we can infer from 'γ is the trajectory of some test particle' to 'γ is a (timelike) geodesic.'

Now, I do not want to suggest that this is a gap in Jacobs' argument—Jacobs is clear about the fact that he is interested in a specific motivationalist worry, which is, as a matter of fact, generally cashed out in terms of a demand for a metaphysically perspicuous characterisation of the common ontology of SRMs. However, Jacobs' argument does make it seem as if the value of intrinsic formalisms stands or falls with whether or not one finds this worry compelling. I think that this is a mistake: intrinsic formalisms have a special value regardless of what one thinks about the role that considerations of ontology have to play in interpreting theories. My aim in this article is to get clear on exactly what this value is.

I do this by drawing out a contrast between the kinds of mathematical expressions that may be interpreted as physically meaningful in an intrinsic vs. an extrinsic formalism: in an intrinsic, but not an extrinsic formalism, are facts about whether some equation may be interpreted as physically meaningful essentially trivial—decidable by inspection, as a simple matter of the mathematical operations used to construct it themselves being well-defined. This means that intrinsic formalisms are better suited for getting a handle on what a theory says about the world, because to find out which equations may be interpreted as saying something physically meaningful, we need only take a cursory look at the form of those equations.

In a little more detail, the structure of this article will be as follows. First, in §2, I introduce some of the distinctions that will be useful to us—between intrinsic and extrinsic formalisms, and between reduction, internal, and external sophistication—and explain how these fit together. In §3, I use these distinctions to raise a problem for Jacobs' argument that extrinsic formalisms are not metaphysically perspicuous. Indeed, I will argue, once we are clear about the fact that (a) extrinsic formalisms come with a restriction on what equations are meaningful within the formalism, and (b) unlike external sophistication, the extrinsic approach does not need to appeal to the idea that SRMs represent the same physical state of affairs to justify this restriction, many of Jacobs' worries about extrinsic formalisms disappear. This takes us to §4, in which I develop this idea to make precise the sense in which some equation being meaningful within an extrinsic formalism is non-trivial, and contrast this with the sense in which this is trivial within an intrinsic formalism. This will allow me to say why it is that intrinsic formalisms are more perspicuous than extrinsic formalisms. §5 concludes.

2 Preliminaries: symmetries and the interpretation of symmetry-related models

In general, a theory has kinematically possible models (KPMs) of the form $\langle \mathcal{V}_1, \mathcal{V}_2, ..., Q_1, Q_2, ...\rangle$, where the $\mathcal{V}_i$ are a collection of (structured) value spaces with domains V_i, and the Q_i are a

[5]Note that at times, Jacobs also talks about interpreting a theory as a matter of getting a handle on both its ontology and ideology, although his discussion is focused mostly on the former. One way to read the arguments of this paper is as saying that Jacobs ought to have also pressed the corresponding worry about being able to 'read off' a theory's ideology from the formalism.

collection of quantities (defined as functions from some $\mathcal{V}_i$ into some $\mathcal{V}_j$). The dynamically possible models (DPMs) of the theory are those KPMs which satisfy the theory's equations of motion. A dynamical symmetry of a theory is a bijection on the KPMs induced by a collection of bijections $\chi_i : V_i \to V_i$ which preserves the space of DPMs.

Faced with (non-isomorphic) SRMs of a theory T, one might wish to reformulate T so that it does not distinguish between SRMs (in the sense that they are either identical or isomorphic). Here, one can distinguish extrinsic and intrinsic approaches to carrying out such a reformulation. On the extrinsic approach, one reformulates T by taking appropriate equivalence classes of symmetry-variant structures in the KPMs of T. In some detail, suppose that $\mathcal{V} = \langle V, \phi \rangle$ is the domain (or codomain) for just one quantity Q (the generalisation to the case of multiple ϕ_i or Q_i is straightforward), and consider a group of G of dynamical symmetries $Q \to \chi^* Q$, $\chi : V \to V$. Then

1. We can construct a new space of models $\langle V, \phi, ..., [Q], ... \rangle$, $Q \in [Q]$ iff $\chi^* Q \in Q$ for all $\chi \in G$. Pairs of SRMs $\langle V, \phi, ..., Q, ... \rangle$, $\langle V, \phi, ..., \chi^* Q, ... \rangle$ now correspond to the same model $\langle V, \phi, ..., [Q], ... \rangle$.

2. When $\chi^* \phi \neq \phi$ for some $\chi \in G$, we can also construct a new space of models $\langle V, [\phi], ..., Q, ... \rangle$, $\phi \in [\phi]$ iff $\chi^* \phi \in [\phi]$ for all $\chi \in G$. Pairs of SRMs $\langle V, \phi, ..., Q, ... \rangle$, $\langle V, \phi, ..., \chi^* Q, ... \rangle$ now correspond to isomorphic pairs of models $\langle V, [\phi], ..., Q, ... \rangle$, $\langle V, [\phi], ..., \chi^* Q, ... \rangle$.

Extrinsic formalisms are all and only those formalisms which are instances of the constructions 1 or 2. An example of 1 is the theory of electromagnetism on Minkowski spacetime, with models $\langle M, \eta_{ab}, [A_a], J^a \rangle$ where the $A_a \in [A_a]$ are all related by exact one-form shifts (i.e. Weatherall's (2016) EM_2'). An example of 2 is Dewar's (2018) Maxwell gravitation, which has models $\langle M, t_a, h^{ab}, [\nabla], T^{ab} \rangle$ where $[\nabla]$ is an equivalence class of rotationally equivalent flat derivative operators. By contrast, the intrinsic approach seeks to reformulate T without appealing to equivalence classes of symmetry-variant structures. Call any formalism which is not an instance of the constructions 1 or 2 an intrinsic formalism. For example, Maxwell gravitation can also be presented as a theory with models $\langle M, t_a, h^{ab}, \circlearrowleft, T^{ab} \rangle$, where $\circlearrowleft$ is a compatible standard of rotation[6] (Chen 2023; March 2023, 2024).

But this is not all there is to an extrinsic formalism. We can get a handle on what else is needed via the following question: what kind of equations is it sensible to write down, when working with an extrinsic formalism? In particular: is it legitimate for these equations to make use of objects in the equivalence classes of symmetry-variant structures?

The answer which I want to suggest, and which I think captures the way in which extrinsic formalisms have previously been thought of, is 'yes'—so long as satisfaction of these equations is independent of the choice of representative of the equivalence class. For example, when Earman defined Maxwellian spacetime as a structure $\langle M, t_a, h^{ab}, [\nabla] \rangle$, he wrote:

> Although questions about the acceleration of a body are not in general meaningful in this setting, it is, of course, meaningful to ask about the state of rotation of a fluid or an extended body. (Earman 1989, p. 32)

And when Dewar (2018) wrote down dynamics for Maxwell gravitation on Earman's Maxwellian spacetime, he went to some trouble to show that satisfaction of these equations was independent of the choice of $\nabla \in [\nabla]$. Or consider: it is often claimed that the only possible field equations for the theory of scalar electrodynamics with models $\langle M, \eta_{ab}, [\langle A_a, \psi \rangle] \rangle$ are ones which are $U(1)$ gauge invariant.[7]

So extrinsic formalisms don't just equivocate between symmetry-variant structures in defining the models of the theory; they also restrict the space of mathematical expression involving these structures which can be 'meaningfully' written down to ones which are invariant under the relevant

[6] This was introduced by Weatherall (2018): if t_a, h^{ab} are compatible temporal and spatial metrics on M, a standard of rotation $\circlearrowleft$ compatible with t_a and h^{ab} is a map from smooth vector fields ξ^a on M to smooth, antisymmetric rank-$(2,0)$ tensor fields $\circlearrowleft^b \xi^a$ on M, such that (i) $\circlearrowleft$ commutes with addition of smooth vector fields; (ii) given any smooth vector field ξ^a and smooth scalar field α, $\circlearrowleft^a (\alpha \xi^b) = \alpha \circlearrowleft^a \xi^b + \xi^{[b} d^{a]} \alpha$; (iii) $\circlearrowleft$ commutes with index substitution; (iv) given any smooth vector field ξ^a, if $d_a(\xi^n t_n) = 0$ then $\circlearrowleft^a \xi^b$ is spacelike in both indices; and (v) given any smooth spacelike vector field σ^a, $\circlearrowleft^a \sigma^b = D^{[a} \sigma^{b]}$.

[7] cf. also Mundy (1986).

class of symmetry transformations.[8] Conversely, equations which are not so invariant—like the acceleration of a fluid relative to some $\nabla \in [\nabla]$ in Maxwellian spacetime—are not to be regarded as 'meaningful.' One way to spell this out is in terms of a supervaluationist semantics, in which the supervaluation is carried out over all objects in the relevant equivalence classes (see e.g. Dewar (2019) and Jacobs (2021a)).

The point I want to press here is that from the perspective of an extrinsic formalism, one does *not* need to appeal to the idea that SRMs represent the same physical state of affairs to justify the claim that equations which are not independent of the choice of $\phi \in [\phi]$ are physically meaningless. For suppose we take seriously the project of characterising structures in the models of a theory T extrinsically, i.e. via a preferred equivalence class of representations $[\phi]$. Then what it is for the models of T to have this kind of structure is *just is* to say: facts about the models of T can be represented equally (but redundantly) by any one of the $\phi \in [\phi]$. Since the models of T are supposed to represent physical states of affairs, it follows that equations which are not independent of the choice of $\phi \in [\phi]$ cannot be interpreted as saying something physically meaningful (since they cannot sensibly be thought of as *about* the models of T).[9] Conversely, insofar as it does make sense to talk about physical facts which are not so independent within the formalism of T, T simply fails to have the kind of structure we have defined it to have.[10] It is for this reason that equations which are not independent of the choice of $\phi \in [\phi]$ are physically meaningless.

That said, I do want to point out that just restricting to equations which are independent of the choice of representative of the equivalence class can't quite be the full story about which equations are meaningful in an extrinsic formalism, since it too easily falls prey to what one might think of as 'spurious' invariances. This is nicely illustrated with the example of Earman's Maxwellian spacetime. Since all the $\nabla \in [\nabla]$ are flat, the equation $R^a{}_{bcd} = 0$ is invariant between objects in the equivalence class, and will come out as being true. But this is, intuitively speaking, the wrong result: Maxwellian spacetime lacks full affine structure, and so it is simply not sensible to speak of it as flat or non-flat (though one can make sense of a weaker notion of rotational flatness, see March (2024)). Probably the right thing to do in this case is to note that one can also represent the rotation standard with a non-flat connection (any connection satisfying $R^{ab}{}_{cd} = 0$ will do), so that the equation $R^a{}_{bcd} = 0$ is not invariant between all the connections which can be used to represent the rotation standard. But it is not straightforward how to spell this out in general, at least without some intrinsic characterisation of the structure of interest already to hand.[11] In any case, I take the above arguments to show that being independent of the choice of representative of the equivalence class is a plausible minimal restriction on which equations are physically meaningful within an extrinsic formalism—so going ahead, we can adopt this restriction along with the proviso that it may need to be tightened up later.

With the intrinsic-extrinsic distinction on the table, I will now discuss how this relates to three different approaches to capturing the invariant content of SRMs which have been distinguished in the literature: reduction, internal sophistication, and external sophistication (see e.g. Dewar (2019) and Martens and Read (2021) for clear expositions of these views). Reduction says that faced with SRMs of a theory T, one should reformulate T so that SRMs all map to the same model of the reduced theory. Importantly, this is the case even if the SRMs in question are isomorphic. By contrast, internal and external sophistication both say that if SRMs of T are isomorphic, one may interpret them as physically equivalent via appeal to anti-haecceitism.[12]

[8] One might worry about how this is supposed to apply to variational principles, but I won't discuss this here.

[9] Compare Lewis (1986): "A proposition [read: equation] is about a subject matter [...] if and only if that proposition holds at both or neither of any two worlds [read: mathematical structures] that match perfectly with respect to that subject matter."

[10] I take this to be in the spirit of Belot's (2000) point that writing down equations which require for their formulation structure which the theory does not posit is "arrant knavery"—i.e. certainly not sensible and maybe even incoherent; cf. also Wallace (2019) on coordinate-based approaches and Myrvold (2019) on Earman's SP2. I will discuss Earman's principles more in §3.

[11] The other obvious thing to do here is to point out that the equation $R^a{}_{bcd} = 0$ coming out as true is not quite as bad as it sounds, since in Maxwellian spacetime, the left hand side of this equation lacks an interpretation in terms of parallel transport of timelike vectors along arbitrary (spacelike or timelike) curves. On this kind of view, the equation $R^a{}_{bcd} = 0$ might be true in Maxwellian spacetime, but it would not follow from this that Maxwellian spacetime is 'flat' in the usual sense of the word (flatness would require in addition e.g. a standard of parallel transport for timelike vectors along timelike curves). I think this kind of response can also probably be made to work, but again, it is not completely obvious how to spell this out in general.

[12] Or its analogue for quantities, anti-quidditism—though the difference does not really matter, since anti-

Where these two approaches differ is on their treatment of non-isomorphic SRMs. According to internal sophistication, one must first reformulate T so that SRMs map to isomorphic models of the sophisticated theory before one can appeal to anti-haecceitism to say that these models represent the same physical state of affairs, whereas according to external sophistication, one can 'stipulate' that SRMs are to 'count' as isomorphic, without reformulating the theory. External sophistication is thus very naturally articulated from the theories-as-categories standpoint, in which one can understand stipulating isomorphisms between non-isomorphic models as meaning that one is to add arrows between non-isomorphic models into one's category of models.[13]

For what it's worth, I also think that external sophistication *should* be understood category-theoretically, if it is not either to be altogether mysterious or to collapse into a variant of reduction or internal sophistication (though one might also take this to mean that the distinction between external sophistication and internal sophistication or reduction was never the relevant one to start off with).[14] For example, Dewar characterises external sophistication as "declaring, by fiat, that the symmetry transformations are now going to 'count' as isomorphisms" (Dewar 2019, pp. 502-503). But as Jacobs (2022) notes, this way of putting it is "somewhat puzzling" (Martens and Read (2021, p. 340) go further, saying that "[to] stipulate that qualitatively distinct, i.e. non-isomorphic models [...] are nevertheless isomorphic reads *prima facie* as nothing more than a flat-out contradiction"). It is in this vein that Jacobs (2021b, 2022) offers his own take on external sophistication, as meaning that the invariant content of SRMs should be captured using an extrinsic formalism. But if it is the provision of an extrinsic formalism which external sophistication is really about, then it becomes clear that it is really a form of reduction or internal sophistication, since it involves mathematically reformulating the models of the theory in such a way that SRMs will end up being either identical or isomorphic. Conversely, if we take Jacobs to mean that external sophistication should be understood as a commitment to characterising structures in the models of a theory extrinsically via their isomorphisms, but without reformulating the theory so that SRMs are in fact isomorphic, then we are back to the worry about how to make sense of 'declared isomorphisms' between non-isomorphic models, i.e. back to square one.

Having said this, it should be clear what I want to say about how the intrinsic vs. extrinsic formalisms distinction relates to the reduction vs. internal sophistication vs. external sophistication distinction, which is that the two are basically orthogonal. To be sure, one area of overlap remains, which is that it is unclear why proponents of external sophistication would ever be interested in the intrinsic vs. extrinsic formalisms distinction, since they can always make do with the formalism they already have. But one can externally sophisticate an intrinsic or extrinsic formalism; likewise, one can reduce a theory by moving to an intrinsic or extrinsic formalism, and one can also internally sophisticate by moving to an intrinsic or extrinsic formalism.

To make this last point absolutely clear, it is helpful to consider an example. Take the theory of Newtonian point-particle mechanics with models $\langle \mathcal{M}, \phi, \mathcal{B}, \gamma(i), \mathbb{R}^+, m(i)\rangle$, where $\mathcal{M} = \langle M, t_a, h^{ab}, \nabla, \xi^a\rangle$, ϕ is a scalar field which represents the gravitational potential, $\mathcal{B}$ is a (structured) domain of particles, the $\gamma(i) : \mathcal{B} \times \mathbb{R} \to \mathcal{M}$, are a collection of (smooth, future-directed) timelike curves which represent particle worldlines, and $m(i) : \mathcal{B} \to \mathbb{R}^+$ is an assignment of mass values to each particle. Suppose that uniform mass scalings—transformations of the form $m(i) \to \psi \circ m(i)$, where ψ is a bijection on the domain R^+ of $\mathbb{R}^+$ which preserves the relation $\leq$ and the operation $+$[15]—are dynamical symmetries of this theory. Table 1 outlines four possible approaches to reformulating the models of the theory, in light of this symmetry.

The first row of table 1 is fairly straightforward. In the intrinsic reduction corner, we replace the assignment of mass values with an assignment of mass ratios $m(i, j) : \mathcal{B} \times \mathcal{B} \to \mathbb{R}^+$ to each pair of particles, subject to the constraints $m(i, i) = 1$ and $m(i, j)m(j, k) = m(i, k)$. In the extrinsic reduction corner, we replace $m(i)$ with an equivalence class of mass value assignments, i.e. $m(i) \in [m(i)]$ iff $\psi \circ m(i) \in [m(i)]$ for all uniform mass scalings ψ. Both of these count as

quidditism about (the determinate magnitudes of) some physical quantity amounts to anti-haecceitism about points in the value space of that quantity.

[13] c.f. Dewar (2019, 2022)

[14] This is how I am sometimes inclined to read Jacobs (2021b), who instead distinguishes 'structure-first' and 'symmetry-first' approaches to capturing the invariant content of SRMs.

[15] Note that uniform mass scalings won't preserve the relation $\times$ on R^+, i.e. they are not automorphisms of the mass value space $\mathbb{R}^+ = \langle R^+, \leq, +, \times\rangle$. In particular, this means that models related by a uniform mass scaling are not isomorphic.

	Intrinsic	Extrinsic
Reduction	$\langle \mathcal{M}, \phi, \mathcal{B}, \gamma(i), \mathbb{R}^+, m(i,j) \rangle$	$\langle \mathcal{M}, \phi, \mathcal{B}, \gamma(i), \mathbb{R}^+, [m(i)] \rangle$
Internal sophistication	$\langle \mathcal{M}, \phi, \mathcal{B}, \gamma(i), D_m, \leq, \circ, m(i) \rangle$	$\langle \mathcal{M}, \phi, \mathcal{B}, \gamma(i), D_m, [f], \mathbb{R}^+, m(i) \rangle$

Table 1: Different reformulations of Newtonian point-particle mechanics, organised by where they fall with respect to the intrinsic vs. extrinsic formalisms and reduction vs. internal sophistication distinctions.

reduced theories: the $m(i,j)$ are invariant under uniform mass scalings, and by construction any pair of models $\langle \mathcal{M}, \phi, \mathcal{B}, \gamma(i), \mathbb{R}^+, m(i) \rangle$, $\langle \mathcal{M}, \phi, \mathcal{B}, \gamma(i), \mathbb{R}^+, \psi \circ m(i) \rangle$ related by a uniform mass scaling map to the same model $\langle \mathcal{M}, \phi, \mathcal{B}, \gamma(i), \mathbb{R}^+, [m(i)] \rangle$.

In the second row of table 1, we have modified the definition of the mass value space. In the intrinsic internal sophistication corner, we have replaced $\mathbb{R}^+$, with an additive extensive structure $\langle D_m, \leq, \circ \rangle$, where D_m is a domain of cardinality $2^{\aleph_0}$, $\leq$ is a total order on D_m, and $\circ$ is an associative binary operation (representing addition of mass values), subject to certain axioms.[16] Since uniform mass scalings are automorphisms of $\langle D_m, \leq, \circ \rangle$, models related by a uniform mass scaling are now isomorphic. In the extrinsic internal sophistication corner, the mass value space $\langle D_m, [f], \mathbb{R}^+ \rangle$ is again an additive extensive structure, but this time we have characterised it extrinsically rather than intrinsically. $[f]$ is an equivalence class of bijections $f : D_m \to \mathbb{R}^+$ defined as follows: $f \in [f]$ iff $f \circ \psi \in [f]$ for any bijection ψ on R^+ which preserves $\leq$ and $+$. To see that this is indeed an instance of internal sophistication, consider any bijection $\psi : D_m \to D_m$ which preserves $[f]$. It follows immediately from the definition of $[f]$ that these bijections are in one-to-one correspondence with our original uniform mass scalings. Now, in general, ψ is not an automorphism of $\langle \mathcal{M}, \phi, \mathcal{B}, \gamma(i), D_m, [f], \mathbb{R}^+, m(i) \rangle$, since $\psi \circ m(i) \neq m(i)$ unless $\psi = \mathrm{id}_{D_m}$, so this is not a reduced formalism. But it is an internally sophisticated formalism: ψ induces an isomorphism of $\langle \mathcal{M}, \phi, \mathcal{B}, \gamma(i), D_m, [f], \mathbb{R}^+, m(i) \rangle$, since it is an automorphism of $\langle D_m, [f], \mathbb{R}^+ \rangle$. This makes my point: that the intrinsic vs. extrinsic formalisms distinction is independent of the reduction vs. internal sophistication distinction.

3 Metaphysical perspicuity

We can now return to Jacobs' argument that only intrinsic formalisms are metaphysically perspicuous. To reiterate, according to Jacobs, the ontological commitments of intrinsic formalisms can simply be 'read off' from the formalism, whereas extrinsic formalisms have only an 'effective decision procedure' for whether the theory is committed to some piece of structure or other, by determining whether that structure can be invariantly defined from objects in the equivalence classes.

The fact that the extrinsic approach has only an effective decision procedure for determining a theory's ontological commitments, Jacobs claims, means that extrinsic formalisms are not metaphysically perspicuous. This is for three reasons. First, extrinsic formalisms limit attempts at causal explanation, since only symmetry-invariant structures can be dynamically efficacious. Second, it is unclear what grounds (or explains, or justifies) the physical equivalence of SRMs from the perspective of an extrinsic formalism. This is supposed to be because the extrinsic approach reverses the natural order of explanation—the theory is committed to a certain ontology because SRMs represent the same physical state of affairs, rather than *vice versa*. Third, extrinsic formalisms appeal to physically irrelevant (i.e. symmetry-variant) quantities to characterise the structure of the theory's value spaces. This does not seem to tell us what these value spaces are really like, even if it fixes the correct structure (this is the constructivist complaint).[17]

[16] See e.g. Hölder (1901) and Krantz et al. (1971) for details of these axioms. In effect, this amounts to 'forgetting' the multiplication operation $\times$ on R^+.

[17] Note that for this kind of reason, extrinsic formalisms are likely to be repugnant to fans of Reichenbachian constructivism (see Adlam et al. (2022) and Linnemann and Read (2021)), though here the issue is not so much that the quantities are 'physically irrelevant,' but rather a problem of not being able to help oneself to structure that one hasn't constructed yet.

However, I think that the extrinsic approach has the resources to resist these worries, at least when we take into account the fact that (a) the extrinsic approach comes with a restriction on what equations are physically meaningful, and (b) the extrinsic approach does not need to appeal to the idea that SRMs represent the same physical state of affairs to justify this restriction (recall §2). Indeed, I think this gives the extrinsic approach a very natural response to Jacobs' concern about what explains the physical equivalence of SRMs. That is, SRMs represent the same physical state of affairs because the only claims which would possibly allow one to distinguish between them (up to isomorphism) are physically meaningless, since they depend on the choice of representative in the equivalence classes. And the fact that the theory is committed to a certain ontology because certain distinctions are physically meaningless, and others physically meaningful, strikes me as the right result.[18] In fact, it seems to me precisely the kind of reasoning involved in Earman's (1989) famous symmetry principles:

SP1: Every dynamical symmetry of T is a spacetime symmetry of T.

SP2: Every spacetime symmetry of T is a dynamical symmetry of T.

In effect, SP2 says that T should posit enough spacetime structure that it can distinguish (in the sense that they are not isomorphic) between models which are not related by a dynamical symmetry. Conversely, SP1 says that T should not posit so much spacetime structure that it distinguishes (again, in the sense that they are not isomorphic) between models which are related by a dynamical symmetry. So providing we hold the dynamical symmetries of T fixed, Earman's principles suggest that it is entirely appropriate to let questions of what distinctions are physically meaningful underwrite questions of a theory's ontology in this way.

Turning now to the causal explanations worry, I think this is misplaced. Granted, the causal explanations which one can read off from an extrinsic formalism will often involve making reference to symmetry-variant quantities. But so long as the explanations themselves are symmetry-invariant,[19] it is not clear why this should hamper attempts at causal explanation. For example, Jacobs claims that the proponent of the extrinsic approach cannot without further argumentation explain the Aharonov-Bohm effect, in which a charged particle in the vicinity of an impenetrable solenoid picks up a phase proportional to the flux through that solenoid. This is supposed to be because the causal story involved in the explanation of the Aharonov-Bohm effect must appeal to invariant structures, such as the holonomies of the electromagnetic one-form. But here is an explanation to which the extrinsicalist can perfectly well appeal: the charged particle picks up a phase because up to $U(1)$ gauge symmetry, the electromagentic one-form can be represented as having a value such-and-such in the region through which the particle travels, and the phase difference picked up by the particle follows from this plus the dynamics of the theory. To the worry that such an explanation is not appropriately *causal* I say: whatever one's favourite account of causation is, either this explanation counts as appropriately causal (as for e.g. counterfactual, interventionist, or productive accounts), or this is a problem for causal explanations of the Aharonov-Bohm effect in general, rather than the explanation which the extrinsic approach offers in particular (as for e.g. conserved quantity or causal mechanisms approaches), see Earman (2024) for recent discussion.

Finally, consider the worry that extrinsic formalisms appeal to physically irrelevant (i.e. symmetry-variant) quantities to characterise the theory's value spaces. Again, I don't think the proponent of the extrinsic approach should find this worry compelling. To begin with, the fact that some quantity is symmetry-variant does not by itself mean that it is physically irrelevant—for some of the degrees of freedom of that quantity may be symmetry-invariant, and thus dynamically efficacious. This is the case for, e.g. the rotational degrees of freedom of the connections in Earman's Maxwellian spacetime, or the inexact degrees of freedom of the electromagnetic one-form. Secondly, the example Jacobs discusses—in which a preferred class of coordinatisations are used to

[18] Compare Leibniz's famous 'shift' argument against the reality of absolute space: absolute space is unreal because "[to] say that God can cause the whole universe to move forwards in a right line, or in any other line, without making otherwise any alteration in it; is another chimerical supposition. For two states indiscernible from each other, are the same state; and consequently, 'tis a change without any change." (Leibniz and Clarke 1998, p. 38)

[19] Note that we do also appear to give symmetry-variant explanations in physics, e.g. the appeal to the rest frame of the rocket in the explanation of Bell's rockets thought experiment, though one might dispute whether this really counts as a symmetry-variant explanation, given that the rest frame of the rocket can be defined invariantly from its worldline. However, this will depend on whether or not one has an operational understanding of the coordinate systems in question.

characterise Galilean spacetime—tends to obscure this point. Of course Jacobs is absolutely right when he says that coordinates themselves are not dynamically efficacious, but this is not the issue. A preferred class of coordinatisations can still be physically relevant because—and this is what Jacobs does not say—those preferred coordinatisations are used to express the dynamics of the theory, and so encode information about the symmetries of those dynamics. And the symmetries of the dynamics are physically relevant precisely because they restrict what kind of dynamics those could be.

As far as Jacobs' more general point goes—that extrinsic formalisms don't tell us what a theory's value spaces are really like—the extrinsic approach has an answer to this too. The answer is that those value spaces have just enough structure to capture all the invariant degrees of freedom of the symmetry-variant quantities that are used to define them, and no more. The proponent of the intrinsic approach will press the question: but just what structures are those? But here I think that the extrinsicalist can simply dig their heels in. Granted, the extrinsicalist cannot say without further argumentation just what the invariant structures in question are, but to insist that they have not said what the value spaces of the theory are really like seems question-begging.

With that said, I do think that Jacobs' insight about the fact that one cannot 'read off' the ontological commitments of extrinsic formalisms is basically on the right track. However, as I have argued, it it not clear why this fact by itself should act as a barrier to metaphysical perspicuity. So let us see what kind of consequences would act as such a barrier.

4 Triviality and non-triviality

Throughout this article, I have pressed the idea that within an extrinsic formalism, equations are not independent of the choice of representative of the equivalence classes of symmetry-variant structures should not be thought of as physically meaningful. I have also argued that from the perspective of an extrinsic formalism, this restriction is well-motivated, and perhaps even compulsory.

But it is also non-trivial. To see this, consider the kinds of equations that can be interpreted as physically meaningful in an intrinsic formalism. Since intrinsic formalisms do not equivocate between symmetry-variant structures in defining the models of the theory, any equation which is (i) constructed out of objects in the models of the theory, and (ii) is mathematically well-defined, expresses a statement which can be true or false of the models of the theory, and so can be interpreted as saying something physically meaningful. In other words, to check whether an equation is physically meaningful within an intrinsic formalism, one only needs to take a cursory look at the form of that equation.

Contrast this with the process of checking whether an equation is physically meaningful within an extrinsic formalism. In this case, not only does one need to verify that the equation is mathematically well-defined—in general, one also needs to verify that satisfaction of that equation is independent of the choice of representative of the equivalence class. And this second step is generally highly non-trivial (think e.g. of Dewar's (2018) dynamics for Maxwell gravitation, or of the many pages of ink that physics undergraduates have spilled over the years verifying $U(1)$ gauge invariance of the Lagrangian of scalar electrodynamics).

But if this is right, we now have a handle on why it is that intrinsic formalisms are more perspicuous than intrinsic formalism. The issue is not just that one cannot read off the theory's ontology from the formalism—though that may come into it too—but that one also cannot read off which equations that can be constructed in that formalism are physically meaningful. This means that intrinsic formalisms are better suited for getting a handle on what a theory says the world is like. For example, if we are interested in cataloguing the kind of things the theory is able to say about the world, then intrinsic formalisms allow us to read off which mathematical expressions may be interpreted as saying something about the world. Of course, those mathematical expressions will still need interpreting, but one does not need to do substantial mathematical heavy lifting in order to decide which bits of uninterpreted mathematics are candidates for being given a physical interpretation. Or, if we are interested in cataloguing what kind of inferences the theory licences, then we will first need to catalogue what kind of equations are candidates to enter into such inferences—intrinsic formalisms allow us to read off what equations these are.

5 Close

In this article, my aim has been to get clear on the special value of intrinsic formalisms, in the context of reformulating a theory with SRMs. I have argued that this value comes from the fact that whether some equation may be interpreted as physically meaningful in an intrinsic formalism is trivial in a way in which it is not for extrinsic formalisms. The special value of an intrinsic formalism is just that it allows us to 'read off' what these equations are.

Of course, being able to 'read off' a theory's ontology is also valuable, if one thinks that considerations of ontology have a privileged role to play in theory interpretation. But my focus on being able to read off what equations are meaningful means that my argument about the value of intrinsic formalisms is able to stand apart from this issue. Rather than just being the purview of a certain brand of philosopher of physics, intrinsic formalisms should be of value to everyone.

Acknowledgements

I would like to thank Dominik Ehrenfels, Caspar Jacobs, and James Read for helpful discussions and comments. I am grateful to Balliol College, Oxford, and the Faculty of Philosophy, University of Oxford, for financial support.

References

Adlam, Emily, Niels Linnemann, and James Read (2022). *Constructive Axiomatics in Spacetime Physics Part II: Constructive Axiomatics in Context*. URL: https://arxiv.org/abs/2211.05672.

Belot, Gordon (2000). "Geometry and Motion". *The British journal for the philosophy of science* 51, pp. 561–595.

Chen, Elliott (2023). "Newtonian gravitation on Maxwell spacetime". [Forthcoming]. URL: http://philsci-archive.pitt.edu/id/eprint/22440.

Dewar, Neil (2018). "Maxwell gravitation". *Philosophy of Science* 85.2, pp. 249–270.

Dewar, Neil (2019). "Sophistication about Symmetries". *The British Journal for the Philosophy of Science* 70.2, pp. 485–521. DOI: 10.1093/bjps/axx021. URL: https://doi.org/10.1093/bjps/axx021.

Dewar, Neil (2022). *Structure and Equivalence*. United Kingdom: Cambridge University Press.

Earman, John (1989). *World Enough and Space-Time*. MIT Press.

Earman, John (2024). *Explaining the Aharonov-Bohm Effect*. URL: https://philsci-archive.pitt.edu/22970/.

Field, Hartry (2016). *Science without Numbers*. 2nd ed. Oxford: OUP.

Hölder, Otto (1901). *Die Axiome der Quantität und die Lehre vom Mass*. Teubner.

Jacobs, Caspar (2021a). "Invariance or equivalence: a tale of two principles". *Synthese* 199.3-4, pp. 9337–9357.

Jacobs, Caspar (2021b). "Symmetries as a Guide to the Structure of Physical Quantities". DPhil thesis. University of Oxford.

Jacobs, Caspar (2022). "Invariance, intrinsicality, and perspicuity". *Synthese* 200.2.

Krantz, David et al. (1971). *Foundations of Measurement*. Mineola, New York: Dover Publications.

Leibniz, Gottfried Wilhelm and Samuel Clarke (1998). *The Leibniz-Clarke correspondence*. Ed. by Henry Gavin Alexander. Manchester: Manchester University Press.

Lewis, David (1986). "A Subjectivist's Guide to Objective Chance". In: *Philosophical Papers: Volume II*. Oxford: OUP, pp. 83–132. DOI: 10.1093/0195036468.003.0004.

Linnemann, Niels and James Read (2021). *Constructive Axiomatics in Spacetime Physics Part I: Walkthrough to the Ehlers-Pirani-Schild Axiomatisation*. DOI: 10.48550/ARXIV.2112.14063. URL: https://arxiv.org/abs/2112.14063.

Luc, Joanna (2023). "Motivationalism vs. interpretationalism about symmetries: some options overlooked in the debate about the relationship between symmetries and physical equivalence". *European journal for philosophy of science* 13.3.

March, Eleanor (2023). *Maxwell gravitation without reference to equivalence classes of derivative operators*. URL: https://philsci-archive.pitt.edu/22404/.

March, Eleanor (2024). *On coordinate-based and coordinate-free approaches to Newtonian gravitation on Maxwellian spacetime*. URL: https://philsci-archive.pitt.edu/22958/.

Martens, Niels C. M. and James Read (2021). "Sophistry about symmetries?" *Synthese* 199.1, pp. 315–344. DOI: 10.1007/s11229-020-02658-4. URL: https://doi.org/10.1007/s11229-020-02658-4.

Møller-Nielsen, Thomas (2017). "Invariance, Interpretation, and Motivation". *Philosophy of Science* 84.5, pp. 1253–1264.

Mundy, Brent (1986). "On the General Theory of Meaningful Representation". *Synthese* 67.3, pp. 391–437.

Myrvold, Wayne (2019). "How could relativity be anything other than physical?" *Studies in the History and Philosophy of Modern Physics* 67, pp. 137–143.

Wallace, David (2019). "Who's afraid of coordinate systems? An essay on the representation of spacetime structure". *Studies in the History and Philosophy of Modern Physics* 67, pp. 125–136.

Weatherall, James Owen (2016). "Are Newtonian gravity and geometrised Newtonian gravity theoretically equivalent?" *Erkenntnis* 81.5, pp. 1073–1091.

Weatherall, James Owen (2018). "A brief comment on Maxwell(/Newton)[-Huygens] spacetime". *Studies in the History and Philosophy of Modern Physics* 63, pp. 34–38.