Sameness and Separability in Gauge Theories

John Dougherty*†

In the philosophical literature on Yang-Mills theories, field formulations are taken to have more structure and to be local, while curve-based formulations are taken to have less structure and to be nonlocal. I formalize the notion of locality at issue and show that theories with less structure are nonlocal. However, the amount of structure had by some formulation is independent of whether it uses fields or curves. The relevant difference in structure is not a difference in set-theoretic structure. Rather, it is a difference in the structure of the category of models of the theory.

1. Introduction. Much of the philosophical literature on Yang-Mills theories is concerned with the differences between interpretations that represent the state of the world using fields and those that represent the state of the world using properties assigned to curves in space-time. These interpretations are inspired by corresponding mathematical formulations of Yang-Mills theory using bundles and holonomies, respectively. The former class—advocated by Weatherall (2016a), for example—is generally thought to deliver a localized picture of the world but also to involve a kind of "surplus structure" (Redhead 2001). On these interpretations, mathematically unequal but gauge-equivalent configurations correspond to the same physical state of affairs, so there is a representational redundancy. The latter class of interpretations—

*To contact the author, please write to: University California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093; e-mail: jedoughe@ucsd.edu.

†For helpful comments, discussion, and guidance, thanks are owed to Ben Feintzeig, Sarita Rosenstock, Sebastian Speitel, Anncy Thresher, Jim Weatherall, and the University of California, Irvine, philosophy of physics reading group. This work was performed under a collaborative agreement between the University of Illinois at Chicago and the University of Geneva and made possible by grant (10) 56314 from the John Templeton Foundation, and its contents are solely the responsibility of the author and do not necessarily represent the official views of the John Templeton Foundation.

Philosophy of Science, 84 (December 2017) pp. 1189–1201. 0031-8248/2017/8405-0032$10.00

advocated by Belot (1998) and especially Healey (2007), for example—is meant to eliminate this surplus structure at the cost of locality. In these theories, the state of some region does not supervene on the state of its subregions (Myrvold 2011). The interpretive choice between these positions is sometimes presented as a cost-benefit analysis, a trade-off between locality and surplus structure.

This accounting oversimplifies the situation. Every claim above is contested: Lyre (2004) and Wallace (2014) argue that field-theoretic interpretations are nonlocal, and Rosenstock and Weatherall (2016) argue that the two classes of interpretations have the same amount of structure. This calls into question the advantages of each interpretation. If bundle interpretations are nonlocal, then they have no advantages over holonomy interpretations and have unnecessary structure on top of that. But if holonomy interpretations have the same amount of structure, then they have no less structure than bundle representations, and so have whatever surplus the bundle interpretations do, too. Sorting out what is really going on here is made difficult by the fact that, by and large, arguments about locality in this literature are semiformal at best. We should like a precise statement, susceptible to proof, of the locality facts in these theories.

I show that the disagreements over locality and structure result from equivocation. There are two classes each of bundle representations and holonomy representations that differ in their structure and locality facts. One class of representations, call them truncated, has less structure than the other class of untruncated representations. According to the definition of locality given below, truncated Yang-Mills representations are nonlocal while untruncated representations are local. Indeed, the part of the representation that gets lopped off by truncation is just the structure representing the locality of the theory. So the distinction between local and nonlocal theories is orthogonal to the distinction between bundle and holonomy representations: each kind of representation has a local and a nonlocal version. The usual story attributes locality to bundle representations and nonlocality to holonomy representations because it considers only the untruncated bundle representations and only the truncated holonomy representations. When we make precise what we mean by locality, it becomes clear that the relevant feature of the theory is whether it is truncated, not whether it involves bundles or holonomies.

Disambiguating these theories also leads to an interesting general lesson. On the standard view, a physical theory is its set of models. Halvorson (2012) has argued that this conception of theories does not do justice to facts about relationships between theories: it identifies distinct theories and distinguishes alternative formulations of the same theory. It also fails to capture ways that one theory might be a specialization or generalization of another. Halvorson concludes from this that a physical theory cannot solely be its collection of models; instead we must keep track of isomorphisms between these models, in the style of Rosenstock, Barrett, and Weatherall (2015). In what follows, we find that these "external" facts about relationships between theories are not the only thing that the standard view misses out on. Local and nonlocal gauge theories have the same set of models but cannot be the same theory precisely because one is local and the other is not. Articulating the difference between these theories requires appealing to other parts of the mathematical structure—in particular, the sameness structure of the theory. Truncating a theory forgets precisely this sameness structure, simultaneously rendering the theory nonlocal. So a view that takes a physical theory to be its set of models also fails to account for "internal" facts about the theory itself.

2. Spaces of Gauge Configurations. I claimed that "bundle formulation" and "holonomy formulation" are importantly ambiguous. In this section we resolve this ambiguity before considering locality. Briefly, we must choose between modeling gauge-related configurations as equal and modeling them as isomorphic. If we go the former route, the theory is nonlocal whether it is a bundle formulation or a holonomy formulation. If we instead model gauge equivalences as isomorphisms, either kind of formulation is local.

Let X be some space-time region, and consider electromagnetism as a $U(1)$ Yang-Mills theory. What is the configuration space of the gauge fields on X? According to the standard story, there are two classes of answers to this question (Belot 1998; Healey 2007). The first class takes the bundle formulation of Yang-Mills theory as its point of departure. According to this approach, a configuration is specified by a connection on a principal $U(1)$ bundle or by some coordinate representation thereof like a vector potential. The second class of answers takes the configuration space to be a collection of holonomy maps, which assign elements of $U(1)$ to curves in X. Many debates in the philosophical literature are concerned with this level of generality, and arguments for or against one class of formulations confer support to all members of that class.

But as I have presented these formulations, they are incomplete. I have only said how to specify a configuration, not how to determine whether two configurations are the same. Specifying the space of gauge configurations requires choosing when two mathematical representatives correspond to the same physical state of affairs. It is generally thought that there are two choices in the case of Yang-Mills theory. This is incorrect. To clearly lay out the correct menu of choices in a Yang-Mills theory, it helps to speak in terms of groupoids, which make these sameness facts explicit.[1]

1. For an introduction to groupoids, see Brown (1987).

Definition 1. A *groupoid* is a category in which all arrows are isomorphisms.

Much of the philosophical literature on the interpretation of Yang-Mills theories already involves groupoids. For example, Rosenstock and Weatherall (2016) argue that bundle and holonomy interpretations have the same amount of structure, in the sense that a particular groupoid of bundles is equivalent to a particular groupoid of holonomy maps. Weatherall (2016b), too, uses groupoids to analyze relative amounts of structure in gauge theories. More generally, Rosenstock et al. (2015) adopt a point of view on which a scientific theory amounts to a groupoid of models, rather than the set of models considered on the standard semantic view of theories.

I will assume for the rest of this discussion that every collection of mathematical objects is a groupoid. That is, I will take them to generalize sets. A description of a collection requires both a description of the objects in the collection and a description of when two objects are the same (i.e., of objects and isomorphisms). In some cases the isomorphism structure is trivial; these are the cases corresponding to sets. More precisely, say that a groupoid is a set if there is at most one isomorphism between any two of its objects.

Most discussions in the philosophical literature assume that any collection must be a set. So when these discussions encounter a groupoid—such as the groupoid of principal bundles and principal bundle isomorphisms between them—they must squash the isomorphism structure into triviality to produce a set. Call this process truncation.

Definition 2. For any groupoid X, its *truncation* $\|X\|$ is the set of isomorphism classes of X considered as a groupoid.

If a groupoid X is a set (in the sense of the previous paragraph), then it is equivalent to its truncation $\|X\|$. If it is not a set, then there is a natural map $X \to \| X \|$ that sends each object to its isomorphism class, but this map has no inverse. So truncation forgets information about a groupoid, in particular information about its isomorphism structure.

Return to the candidate configuration spaces for a Yang-Mills theory. For simplicity, consider a simply connected space-time region X, and consider a formulation in terms of vector potentials. Let $\Omega^1(X)$ denote the set of all such vector potentials. If we take this set to be the configuration space, then we have a well-known problem: any two vector potentials related by a gauge transformation are empirically and dynamically indistinguishable, leading to empirical underdetermination and dynamical indeterminism. In response, physicists take gauge-related potentials to represent the same physical configuration. That is, the set $\Omega^1(X)$ is too fine because it includes spurious distinctions between gauge-related vector potentials.

We cannot take the set of vector potentials to be the configuration space. Instead we want a space that is obtained from $\Omega^1(X)$ by identifying any two gauge-related configurations. There are two ways to do this using groupoids. If we restrict ourselves to sets, there's only one option: take the set of gauge-equivalence classes of potentials. But groupoids allow for nontrivial sameness structure, and we can take advantage of this. Let $\Omega^1(X) /\!\!/ U(1)$ be the groupoid in which an object is a vector potential and an isomorphism is a gauge transformation. This is a kind of quotient of $\Omega^1(X)$, in that it starts with the set of potentials and then makes any gauge-related configurations the same. But it is not the familiar set-theoretic quotient into equivalence classes. That, rather, is the set $\| \Omega^1(X) /\!\!/ U(1) \|$, the truncation of our weaker quotient.

Assume that gauge-related potentials are the same physical configuration. It remains to decide whether the configuration space is the groupoid where gauge transformations are isomorphisms or the truncation thereof. Relaxing the assumption that our space-time is simply connected, one presentation of the bundle and holonomy formulations of the configuration space is as follows:[2]

> **Definition 3.** Let X be a smooth manifold and G a Lie group with Lie algebra $\mathfrak{g}$. Denote the *groupoid of principal G-bundles with connection on X* by $\mathbf{B}G_{\text{conn}}(X)$, in which an object is a principal G-bundle on X equipped with a principal G-connection, and an isomorphism is a principal bundle isomorphism that restricts to the identity on the base space and preserves the connection.

> **Definition 4.** Let X be a smooth manifold and G a Lie group. Denote the *groupoid of G-valued holonomy maps on X* by $[\mathbf{P}_1(X),\mathbf{B}G]$, in which an object is a map from the path groupoid of X to the group G considered as a one-object groupoid, and an isomorphism is an isomorphism of such functors.

Both of these definitions are underspecified, but we are concerned with a high-level distinction between bundle and holonomy formulations, so this is fine. It will suffice for our discussion to keep in mind the simpler electromagnetic case of the groupoid $\Omega^1(X) /\!\!/ U(1)$, where an object is a vector potential and an isomorphism is a gauge transformation. When X is simply connected, this groupoid is equivalent to $\mathbf{B}G_{\text{conn}}(X)$.

Neither of the groupoids just defined is a set, and so neither is equivalent to its truncation. So we seem to have four candidate configuration spaces: the

2. These definitions are taken from Schreiber (2013, sec. 1.2.6.1.1); see there for more rigorous statements.

bundle formulation $\mathbf{B}G_{\mathrm{conn}}(X)$, the holonomy formulation $[\mathbf{P}_1(X),\mathbf{B}G]$, and the truncations of each. However, it follows from Rosenstock and Weatherall's (2016) main result that $\mathbf{B}G_{\mathrm{conn}}(X)$ and $[\mathbf{P}_1(X),\mathbf{B}G]$ are equivalent, so they are two presentations of the same groupoid. Hence, their truncations are also equivalent. As such, there are really only two options: the groupoid $\mathbf{B}G_{\mathrm{conn}}(X)$ and its truncation $\|\mathbf{B}G_{\mathrm{conn}}(X)\|$. And these really are different because the former is not a set and the latter is.

This menu of spaces resolves one of the tensions discussed in the opening. The standard story has it that bundles involve more structure than holonomy models, but Rosenstock and Weatherall show that there is a precise sense in which a particular bundle representation has the same amount of structure as a particular holonomy representation. As we now see, the standard story rests on an ambiguity. By the measure of structure Rosenstock and Weatherall use, $\mathbf{B}G_{\mathrm{conn}}(X)$ and $[\mathbf{P}_1(X),\mathbf{B}G]$ have the same amount of structure, and their truncations have the same amount of structure, but $\mathbf{B}G_{\mathrm{conn}}(X)$ and $[\mathbf{P}_1(X),\mathbf{B}G]$ have more structure than their truncations. In the standard story, "the" bundle representation is usually taken to be $\mathbf{B}G_{\mathrm{conn}}(X)$, and "the" holonomy representation is usually taken to be $\|[\mathbf{P}_1(X),\mathbf{B}G]\|$. So, properly interpreted, this story is right. It is just not the whole story. Worse, this terminology has not be consistently maintained. Wu and Yang, for example, do not distinguish between $[\mathbf{P}_1(X),\mathbf{B}G]$ and its truncation, referring to the truncated state space as merely "less easy to use" (1975, 3846) than the untruncated one. But the details of their argument rely on structure lost in truncation, so this is a matter of substance, not convenience. A defense of a holonomy interpretation based on the truncated state space's having less structure must show that the truncated state space has all the resources Yang-Mills theory requires, so it cannot appeal to an analysis like Wu and Yang's, which does not attend to the distinction.

Turn to the second tension, concerning locality.

3. Locality, Functorially. The particular notion of locality that interests us is separability. A physical theory is separable if the physical state of some region supervenes on the physical states of its subregions. To make this precise, we need a way of spelling out how the mereological structure of space-time is reflected in the structure of the space of possible states of regions. For any field theory, the state of a region involves at least the states of its subregions. In particular, there is a duality between parthood and determination: if U is a subregion of X, then a physical state of X induces a physical state of U when we restrict our attention. The assignment of configuration spaces to space-time regions is thus functorial—it respects the composition structure of space-time. If the theory is separable, then the state of a region also involves no more than the states of its subregions. So we will formalize separability as a property of the functor that assigns configuration spaces to regions of space-time. Informally, this functor is separable if the configuration space it assigns to some region is the same as the space of collections of states of its subregions.

Einstein formulated the earliest version of the principle of separability as an articulation of a difference between classical and quantum theories. Broadly, a theory is separable if the state it assigns to a system is determined by the state assigned to its subsystems. For a classical field theory, these systems are regions of space-time, with subregions as subsystems. Einstein's principle can be sharpened up into a more formal criterion, versions of which have been given by Belot (1998, 544), Healey (2007, 125), and Myrvold (2011). Following Myrvold, take some manifold X and consider covers of X by open sets. A cover $\mathfrak{V}$ of X is finer than a cover $\mathfrak{U}$ if every region in $\mathfrak{V}$ is a subregion of some region in $\mathfrak{U}$. We can give a semiformal definition of separability as follows:[3]

> For any space-time region X, there are arbitrarily fine open covers $\mathfrak{U}$ of X such that the state of X supervenes on the states of the elements of $\mathfrak{U}$.

As these authors have pointed out, the Aharonov-Bohm (AB) effect (Aharonov and Bohm 1959) demonstrates a failure of separability for holonomy formulations. If $\{U, V\}$ is a cover of the exterior region of an infinite solenoid containing magnetic flux Φ by simply connected regions, as in figure 1, then there is only one possible state for each of U and V, represented by the trivial holonomy. But the state of $U \cup V$ depends on the current in the solenoid, and there are infinitely many different ways that $U \cup V$ could be. So a difference in the state $U \cup V$ does not imply a difference in the states of U or V.

However, this criterion is still ambiguous. It formalizes the mereological structure of space-time, but the notion of supervenience remains intuitive. To formalize the claim that the states of the subregions determine the states of the entire region, we need to model the way that possibilities for subregions interact with the mereological structure of space-time. For a field theory like Yang-Mills theory, the state of a region determines the state of its subregions: if you tell me the electromagnetic facts in this building, I can tell you the electromagnetic facts in each room. More formally, a Yang-Mills theory has a presheaf of configuration spaces:[4]

3. This analysis is adapted from Myrvold (2011, 425). See also Belot (1998, 540) and Healey (2007, 46).

4. See Mac Lane and Moerdijk (1992) for an introductory treatment of presheaves of sets, which are a special case of our presheaves.

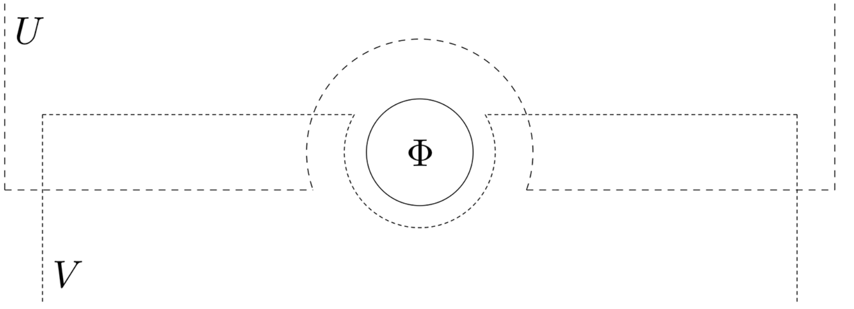

Figure 1. AB setup.

Definition 5. Let X be a topological space. A *presheaf (of groupoids) on X* is a functor $F : \mathcal{O}(X)^{op} \to$ Grpd, where Grpd is the category of groupoids and $\mathcal{O}(X)$ is the category in which an object is an open set $U \subseteq X$ and for any objects V and U there is a unique arrow $V \to U$ just in case $V \subseteq U$.

Presheaves capture the duality remarked on at the start of this section. If V is a subregion of U, then there is an arrow $V \to U$, and a presheaf F sends this to an arrow $F(U) \to F(V)$. If we think of F as an assignment of possibility spaces to regions of space-time, then this map $F(U) \to F(V)$ is just the restriction map.

In physics, a field theory involves a presheaf of configuration spaces on space-time. For example, consider a theory involving some scalar field on a space-time X—a mass density, say, or a gravitational potential. The possible field configurations on X are elements of the set $C^\infty(X, \mathbb{R})$ of real-valued functions on X. If U is a subregion of X, then the possible configurations of U are elements of the set $C^\infty(U, \mathbb{R})$. Given a configuration ϕ on X in $C^\infty(X, \mathbb{R})$, we obtain a configuration $\phi|_U$ in $C^\infty(U, \mathbb{R})$ via restriction. So the presheaf of configurations in this theory is the functor $C^\infty(-, \mathbb{R}) : \mathcal{O}(X)^{op} \to \mathrm{Set}$.

If we think of separability in terms of presheaves, it amounts to the claim that the restriction maps can sometimes be reversed. Consider the AB setup. By restriction, a configuration in $F(U \cup V)$ gives a configuration in $F(U)$ and one in $F(V)$, and these configurations agree on $F(U \cap V)$. So there is a map taking a configuration in $F(U \cup V)$ to a collection of compatible configurations on the cover. If the theory is separable, we can go the other way: a collection of compatible configurations on the cover determines a configuration in $F(U \cup V)$. Moreover, these two maps are inverses. Two different elements of $F(U \cup V)$ are sent to different collections of configurations on the cover because there can be no difference in the state of $U \cup V$ without a difference in the state of some subregion in the cover—this is just what it means for the state of $U \cup V$ to supervene on the state of its subregions. This

https://doi.org/10.1086/694083 Published online by Cambridge University Press

makes the restriction map injective, and since we have a determination map in the other direction it must be a bijection.

Formalizing the compatibility condition on a collection takes a bit of care if it is to be done in a gauge-invariant way. To see how things could go wrong, consider the presheaf of configuration spaces $\Omega^1(-) /\!\!/ U(1)$. Naively, a compatible collection of configurations for the AB setup is a pair of vector potentials A_U on U and A_V on V such that $A_U = A_V$ on the overlap. However, this statement of the compatibility condition is not gauge invariant. If (A_U, A_V) is such a pair (e.g., $A_U = A_V = 0$) and A'_V is gauge equivalent to A_V (e.g., $A'_V = d\theta/2\pi r$), then (A_U, A'_V) will generally not be a compatible collection. But (A_U, A_V) and (A_U, A'_V) are gauge equivalent, because A_V is gauge equivalent to A'_V. So this statement of the compatibility condition distinguishes between isomorphic collections. This makes it gauge variant and ill defined with respect to the groupoid structure. So this cannot be the right statement of the compatibility condition.

The problem we face is familiar to mathematicians, and they have developed a extensive toolkit called abstract homotopy theory to deal with problems like it. In our AB case, the theory says that the groupoid of compatible collections of configurations on the cover is the groupoid where an object is a triple (A_U, A_V, g) of configurations A_U and A_V on U and V, respectively, and a gauge transformation g from A_U to A_V on the overlap $U \cap V$. An isomorphism in this groupoid between objects (A_U, A_V, g) and (A'_U, A'_V, g') is a pair of gauge transformations h and h' between the first two entries of the triple such that $hg' = gh'$.

Generalizing this situation, for any presheaf F on a space X and cover $\mathfrak{U}$ of X we define the collection of compatible configurations using the homotopy limit, which turns a diagram of configuration spaces and restriction maps into a groupoid of collections of configurations that are compatible with respect to all of the restriction maps in an isomorphism-invariant way. If the configuration space $F(X)$ is the same as the space of compatible collections for every cover, then we say that F is separable since it is separable with respect to every cover. In mathematical parlance this means that F is a stack.[5]

Definition 6 (Hollander 2008, def. 1.3). For any topological space X, a presheaf F on X is a *stack on X* if for any good cover $\mathfrak{U}$ of X the natural map

$$F(X) \to \mathrm{holim}\left(\prod_{U \in \mathfrak{U}} F(U) \rightrightarrows \prod_{U, U' \in \mathfrak{U}} F(U \cap U') \substack{\rightarrow\\\rightarrow\\\rightarrow} \dots\right)$$

5. The requirement that the cover be “good” is a technical condition encoding the “sufficiently fine” clause in the semiformal statement above. See Schreiber (2013) for a discussion of its importance.

https://doi.org/10.1086/694083 Published online by Cambridge University Press

is an equivalence. Call this homotopy limit the *groupoid of descent data* for F with respect to $\mathfrak{U}$.

Note that the stack condition is well defined, in that it does not distinguish between equivalent presheaves: if two presheaves are equivalent and one is a stack, then so is the other. Informally, definition 6 says that a presheaf is a stack if a configuration of the region X is the same thing as a compatible collection of configurations of the subregions of X, for any way of carving X up into subregions.

I take the stack condition to be a precise statement of separability. For the concept to apply, the theory must assign configuration spaces to all of the subregions of some space-time region, and the state of the region determines the state of its subregions. So we should be concerned with presheaves. The stack condition then formalizes the idea that the state of the entire region supervenes on the states of its subregions. Any two different collections of compatible subregions determine different configurations of the total region and vice versa. So there cannot be a difference in the configuration of the total region without a difference in some subregion. With this analysis of separability in hand, we can ask whether our Yang-Mills theories are separable.

4. Separability in Gauge Theories. The untruncated groupoids of definitions 3 and 4 generalize to stacks; their truncated cousins do not. In closing, I will sketch proofs of these claims. The arguments from Lyre (2004) and Wallace (2014) are proofs of the latter fact. Myrvold's (2011) proof of the nonseparability of Healey's (2007) holonomy theory is dual to these, showing that the pre-cosheaf assigning algebras of observables to space-time regions is not a costack. Since observables are, broadly speaking, functions on state space, this argument is dual to Lyre's and Wallace's. Benini, Schenkel, and Szabo (2015) give a similarly dual argument that the untruncated algebras of observables are a costack. The fact that the configuration space of a gauge theory is a stack is urged by Schreiber (2013), and the discussion here broadly follows his.

Consider one last time the AB setup. We can take the configuration in the exterior region to be (the principal connection corresponding to) the potential $A = \Phi/2\pi r\, d\theta$, where Φ is the magnetic flux through the solenoid. Suppose that A and A' are the potentials corresponding to fluxes Φ and Φ', respectively. Potentials A and A' are gauge equivalent just in case Φ and Φ' differ by a multiple of 2π; otherwise they specify two different equivalence classes $[A]$ and $[A']$ in the configuration groupoid $\| \mathbf{B}U(1)_{\text{conn}}(U \cup V) \|$. The groupoid of descent data in this case is the set of pairs of gauge-equivalence classes of configurations on U and V that are equal on the overlap. But no matter the value of Φ, the restriction $A|_U$ to U is gauge equivalent to the vanishing potential, and likewise for $A'|_U$. It follows that both $[A]$ and $[A']$ are mapped to the same descent datum, the pair ([0],[0]). Since generally $[A]\neq[A']$, the map appearing in the stack condition is not injective. Hence it is not an equivalence, and we have shown that

Proposition 1. The presheaf $\|\mathbf{B}G_{\mathrm{conn}}(-)\|$ is generally not a stack.

It follows from this proposition and the equivalence between $\mathbf{B}G_{\mathrm{conn}}$ and $[\mathbf{P}_1(-),\mathbf{B}G]$ that the presheaf $\|[\mathbf{P}_1(-),\mathbf{B}G]\|$ is also not a stack. So both of the truncated configuration spaces are nonseparable.

The untruncated configuration spaces avoid this counterexample, and we can prove that they satisfy the stack condition. The configurations in $\mathbf{B}G_{\mathrm{conn}}(U \cup V)$ corresponding to A and A' are not isomorphic unless the difference in their respective fluxes Φ and Φ' is a multiple of 2π. The former is mapped to the descent datum $(A|_U, A|_V, 1)$, and the latter to $(A'|_U, A'|_V, 1)$, where the third entry in both is the gauge transformation corresponding to the constant identity function on $U \cap V$. As before, we can use the fact that both potentials are gauge equivalent to 0 when restricted to U or V to transform them away. The triple corresponding to A becomes $(0, 0, e^{i\lambda})$, and the triple corresponding to A' becomes $(0, 0, e^{i\lambda'})$, where λ is a locally constant real-valued function on $U \cap V$ such that the difference $\lambda|_U - \lambda|_V$ is equal to Φ, and likewise for λ' and Φ'. So the two triples are not the same: the third entry of the triple retains the information about the magnetic flux up to a multiple of 2π. The isomorphism structure of the groupoid allows the compatibility condition to encode the difference between the two configurations, avoiding the counterexample.

More generally, we have

Proposition 2. $\mathbf{B}G_{\mathrm{conn}}$ is a stack.

Proof sketch. Let $\mathfrak{g}$ be the Lie algebra of G, and let $\Omega^1(-;\mathfrak{g}) /\!/ G$ be the presheaf of $\mathfrak{g}$-valued 1-forms weakly quotiented by the action of the gauge group G. To any presheaf F there is associated a stack, the stackification of F, which is the universal way of making F a stack (Laumon and Moret-Bailly 2000, lemma 3.2). The stackification of $\Omega^1(-;\mathfrak{g}) /\!/ G$ is $\mathbf{B}G_{\mathrm{conn}}$; hence, it is a stack. QED

In sum, there are separable and nonseparable bundle formulations, and there are separable and nonseparable holonomy formulations. For each bundle formulation there is an equivalent holonomy formulation, and this equivalence respects separability. The association of holonomy representations with nonseparability rests on the coincidental focus in the literature on $\|[\mathbf{P}_1(-),\mathbf{B}G]\|$ as the paradigmatic holonomy representation.

5. Conclusion. The primary geographic feature in the interpretive landscape of Yang-Mills theories has been a division between bundle formulations and holonomy formulations. Other theoretical features—determinism, empirical underdetermination, locality, and more—have been mapped along this division. In particular, bundle formulations have generally been treated as separable, while holonomy interpretations have not (Healey 2007, sec. 2.4). I have argued above that this neglects an important feature of Yang-Mills theories. The distinction between separable and nonseparable theories is unrelated to the distinction between bundles and holonomies. There are separable theories with equivalent configuration stacks $\mathbf{B}G_{\text{conn}}$ and $[\mathbf{P}_1(-),\mathbf{B}G]$, and there are nonseparable theories with equivalent configuration presheaves $\|\mathbf{B}G_{\text{conn}}(-)\|$ and $\|[\mathbf{P}_1(-),\mathbf{B}G]\|$. So even after choosing between bundle and holonomy formulations, we still must choose between separable and nonseparable versions of these.

The foregoing discussion aimed to point out that there are more choices to make than is usually supposed; it did not try to adjudicate this choice. An argument for the untruncated, separable choice would have to appeal to further assumptions about what we are doing when interpreting classical Yang-Mills theory. If we adopt the cost-benefit analysis of the opening paragraph, then the discussion here leads to an argument for the untruncated theory: it is local, and the "surplus" structure is not surplus after all—it represents locality facts (although the precise way in which it represents these locality facts is unclear). If we are motivated by understanding quantized Yang-Mills theory, we need some story about the correspondence between the quantum and the classical, along with a sense of whether and how differences in separability make a quantum difference. Benini et al. (2015) argue along these lines, claiming that the separable theory is required if we want to get the global algebra of quantum observables right. Schreiber (2013, sec. 1.1.1), too, has argued that nonperturbative effects in quantum field theory require the structure lost in truncation. Arguments along these lines will be pursued elsewhere.

Finally, note that both separability and the amount of structure turn on the difference between a state space and its truncation. This difference is "internal" to the theory, in the sense that it is a fact about the theory itself, not about how the theory stands in relation to other theories or formulations. To be sure, the groupoid structure of $\mathbf{B}G_{\text{conn}}(X)$ makes a difference to its standing with respect to other theories. It means that $\mathbf{B}G_{\text{conn}}(X)$ has the same amount of structure as $[\mathbf{P}_1(X),\mathbf{B}G]$ and more than $\|[\mathbf{P}_1(X),\mathbf{B}G]\|$, for example. But this groupoid structure also makes $\mathbf{B}G_{\text{conn}}$ separable, and this is just a fact about $\mathbf{B}G_{\text{conn}}$. So if we think that separability is a feature of theories—and it is hard to see what else it could be a feature of—then a theory must be more than its underlying set of models.

REFERENCES

Aharonov, Y., and D. Bohm. 1959. “Significance of Electromagnetic Potentials in the Quantum Theory.” *Physical Review* 115:485–91.

Belot, G. 1998. “Understanding Electromagnetism.” *British Journal for the Philosophy of Science* 49 (4): 531–55.

Benini, M., A. Schenkel, and R. Szabo. 2015. “Homotopy Colimits and Global Observables in Abelian Gauge Theory.” *Letters in Mathematical Physics* 105 (9): 1193–1222.

Brown, R. 1987. “From Groups to Groupoids: A Brief Survey.” *Bulletin of the London Mathematical Society* 19 (2): 113–34.

Halvorson, H. 2012. “What Scientific Theories Could Not Be.” *Philosophy of Science* 79 (2): 183–206.

Healey, R. 2007. *Gauging What's Real*. Oxford: Oxford University Press.

Hollander, S. 2008. “A Homotopy Theory for Stacks.” *Israel Journal of Mathematics* 163 (1): 93–124.

Laumon, G., and L. Moret-Bailly. 2000. *Champs algébriques*. Dordrecht: Springer.

Lyre, H. 2004. “Holism and Structuralism in $U(1)$ Gauge Theory.” *Studies in History and Philosophy of Science* B 35 (4): 643–70.

Mac Lane, S., and I. Moerdijk. 1992. *Sheaves in Geometry and Logic*. Dordrecht: Springer.

Myrvold, W. C. 2011. “Nonseparability, Classical, and Quantum.” *British Journal for the Philosophy of Science* 62 (2): 417–32.

Redhead, M. 2001. “The Intelligibility of the Universe.” *Royal Institute of Philosophy Supplement* 48:73–90.

Rosenstock, S., T. W. Barrett, and J. O. Weatherall. 2015. “On Einstein Algebras and Relativistic Spacetimes.” *Studies in History and Philosophy of Science* B 52:309–16.

Rosenstock, S., and J. O. Weatherall. 2016. “A Categorical Equivalence between Generalized Holonomy Maps on a Connected Manifold and Principal Connections on Bundles over That Manifold.” *Journal of Mathematical Physics* 57 (10): 102902.

Schreiber, U. 2013. “Differential Cohomology in a Cohesive ∞-Topos.” Unpublished manuscript, arXiv.org, Cornell University. arXiv:1310.7930v1.

Wallace, D. 2014. “Deflating the Aharonov-Bohm Effect.” Unpublished manuscript, arXiv.org, Cornell University. arXiv:1407.5073.

Weatherall, J. O. 2016a. “Fiber Bundles, Yang-Mills Theory, and General Relativity.” *Synthese* 193 (8): 2389–2425.

———. 2016b. “Understanding Gauge.” *Philosophy of Science* 83 (5): 1039–49.

Wu, T. T., and C. N. Yang. 1975. “Concept of Nonintegrable Phase Factors and Global Formulation of Gauge Fields.” *Physical Review* D 12:3845–57.