Separability in gauge theories*

John Dougherty

November 1, 2016

Abstract

Much of the philosophical literature on classical Yang–Mills theories is concerned with the differences between interpretations that represent the state of the world in terms of fields and those that represent the state of the world in terms of properties assigned to curves in spacetime. In the philosophical literature on Yang–Mills theories, field formulations are taken to have more structure and to be local in a particular sense, while the alternative curve-based formulations are taken to have less structure at the cost of nonlocality. I formalize the notion of locality at issue and show that theories with less structure are indeed nonlocal. However, the amount of structure had by some formulation is independent of whether it uses fields or curves. This leads to a general lesson about structure in mathematized theories. The difference in structure corresponding to the difference in locality is not a difference in set-theoretic structure. Rather, it is a difference in the structure of the collection of models of the theory considered as a category.

Much of the philosophical literature on classical Yang–Mills theories is concerned with the differences between interpretations that represent the state of the world in terms of fields and those that represent the state of the world in terms of properties assigned to curves in spacetime. These interpretations are inspired by corresponding mathematical formulations of Yang–Mills theory: in terms of fiber bundles and holonomies, respectively. The former class—advocated by Arntzenius (2012) and Leeds (1999), for example—are 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—advocated by Belot (1998) and especially Healey (2004, 2007), for example—are 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 (Lyre, 2004).

*Draft in progress; comments more than welcome. For helpful comments, discussion, and guidance, thanks are due to Ben Feintzeig, Sarita Rosenstock, Sebastian Speitel, Anncy Thresher, Jim Weatherall, and the UC Irvine philosophy of physics reading group.

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 (2015) argue that the two classes of interpretations have the same amount of structure. This calls into question the advantages of each interpretation. If fiber 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 fiber bundle representations, and so have whatever surplus the fiber bundle interpretations do, too. Sorting out what's 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 fiber 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 fiber bundle and holonomy representations: each kind of representation has a local and a nonlocal version. The usual story attributes locality to fiber bundle representations and nonlocality to holonomy representations because it considers only the untruncated fiber 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 fiber 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, 2015) 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. 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 separable 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.

1 Separability and functors

The particular notion of locality that concerns 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 spacetime is reflected in the structure of the space of possible states of regions. The state of a region involves at least the states of its subregions; a theory is separable if this is all there is to say. 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 spacetime regions is thus functorial—it respects the composition structure of spacetime. So we will formalize separability as a property of the functor that assigns configuration spaces to regions of spacetime.

Einstein (1935) formulated the earliest version of the principle of separability as an articulation of a difference between classical and quantum theories (Howard, 1985). 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 spacetime, with subregions as subsystems. Einstein took such a theory to be a definitive example of separability, "in that it localizes within infinitely small (four-dimensional) space-elements the elementary things existing independently of one another that it takes as basic" (trans. Howard, 1985, 188). It would be striking if classical Yang–Mills theory also turned out to be nonseparable. In the first place, this would mean that Einstein was wrong to take this principle to distinguish the quantum from the classical. Moreover, the kind of nonseparability at issue here is unrelated to the violation of separability by Bell-type correlations. If the nonseparability arguments are right, then quantum Yang–Mills theory is doubly nonseparable.

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 then give a semiformal definition of separability as follows:[1]

> For any spacetime 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 effect demonstrates a failure of separability for holonomy formulations. If $\{U, V\}$ is a cover of the exterior region of an infinite solenoid by simply-connected regions, 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

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

$U \cup V$ 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 spacetime, but the notion of supervenience remains intuitive. And as we will see, it can be made precise in more than one way. We can begin by spelling out how the possibilities for subregions interact with the mereological structure, further refining the approach of the previous paragraph.

Definition 1. Let X be a topological space. A *presheaf on* X is a functor $F : \mathcal{O}(X)^{\mathrm{op}} \to \mathsf{Set}$, where $\mathcal{O}(X)$ 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 upon 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 spacetime, then this map $F(U) \to F(V)$ is just the restriction map.[2]

In physics, a field theory involves a presheaf of configuration spaces on spacetime. For example, consider a theory involving some scalar field on a spacetime 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

$$C^\infty(-, \mathbb{R}) : \mathcal{O}(X)^{\mathrm{op}} \to \mathsf{Set}$$

More generally, any theory formulated in terms of fiber bundles comes with a presheaf of configurations; a fiber bundle $p : E \to X$ is equivalently a presheaf

$$U \mapsto \Gamma_U E = \{s : U \to E \mid p \circ s = \mathrm{id}_U\}$$

called the presheaf of local sections of the bundle E.[3] Again, any element s of $\Gamma_X E$ gives an element $s|_U$ of $\Gamma_U E$ via restriction. Since a great many geometrical objects in mathematical physics are sections of fiber bundles, presheaves are ubiquitous.

[2] The *locus classicus* for sheaf theory is Mac Lane and Moerdijk (1992).

[3] Actually, it is usually called the sheaf of local sections; see below. The re-expression of a fiber bundle as a sheaf is not possible for fiber bundles equipped with more stuff—for example, a G-structure. This is why fiber bundles are sheaves but principal G-bundles are not: principal G-bundles involve more stuff than plain fiber bundles.

If we think of separability in terms of presheaves, it amounts to the claim that the restriction maps can sometimes be reversed. Consider a cover $\mathfrak{U}$ of a spacetime X and a presheaf F on X. By restriction, a configuration in $F(X)$ gives a configuration in $F(U)$ for each $U \in \mathfrak{U}$. Collecting all of these maps together gives a map

$$F(X) \to \prod_{U \in \mathfrak{U}} F(U)$$

that sends a configuration of X to a collection of configurations, one for each element $U \in \mathfrak{U}$. If this theory is separable, we can go the other way: a collection of assignments of configurations to the elements of $\mathfrak{U}$ determines a configuration in $F(X)$. Moreover, these two maps are inverses. Two different elements of $F(X)$ are sent to different collections in $\prod_{U \in \mathfrak{U}} F(U)$, because there can be no difference in the state of X without a difference in the state of some subregion in the cover—this is just what it means for the state of X to supervene on the state of its subregions. This makes the restriction map injective, and since we have a determination map in the other direction it must be a bijection.

Making precise this invertibility claim involves some subtlety, for two reasons. First, we don't expect *every* element of $\prod_{U \in \mathfrak{U}} F(U)$ to give us an element of $F(X)$. For example, consider any two overlapping $V, U \in \mathfrak{U}$. In our mass density theory above, there is an element of $\prod_{U \in \mathfrak{U}} F(U)$ corresponding to V's being filled with a constant nonzero mass density and U's being empty. This is plainly contradictory: the overlap can't be empty and filled with a nonzero density. We really care about those elements of $\prod_{U \in \mathfrak{U}} F(U)$ corresponding to *compatible* collections of configurations in the regions of $\mathfrak{U}$. But, second, both this notion of compatibility and the notion of invertibility require us to be precise about our notions of sameness. And sameness is complicated in gauge theories, for such a "theory offers (an infinite set of) distinct representations of ostensibly the same physical situation" (Healey, 2007, xvi). Since separability is a property of the theory, not the representation, the relevant notion of sameness is one that tracks physical situations, not mathematical representations. Carefully treating sameness in the context of fiber bundles and holonomies will bring out the fact that both admit of a separable and non-separable formulation.

2 Sheaves

The first of our issues is easier to solve. Some collection of configurations is compatible if the configurations agree wherever the regions overlap. More formally, a pair of configurations in $F(U)$ and $F(V)$ is compatible if they are sent to the same element of $F(U \cap V)$ by the restriction maps. This rules out the mass density example just considered because the two configurations on U and V are sent to different configurations on $U \cap V$. To define separability for a presheaf, we simply need to generalize this to arbitrary collections. This is straightforward, so this section is essentially just a statement of the definition. The next section compares it to the literature on separability of gauge theories.

The semiformal definition of separability from the previous section can be formalized rather cleanly if we introduce some notation. For any cover $\mathfrak{U}$ of X and presheaf F on X, define

$$F(\mathfrak{U}_n) = \prod_{U_1,\ldots,U_n \in \mathfrak{U}} F(U_1 \cap \cdots \cap U_n)$$

So an element A of $F(\mathfrak{U}_1)$ assigns to each element U_i of the cover an element A_i of $F(U_i)$, and an isomorphism $A \cong A'$ in $F(\mathfrak{U}_1)$ is a collection of isomorphisms $A_i \cong A'_i$. An element A of $F(\mathfrak{U}_2)$ takes two regions U_i and U_j of the cover and gives an element A_{ij} of $F(U_i \cap U_j)$, and an isomorphism $A \cong A'$ in $F(\mathfrak{U}_2)$ is a collection of isomorphisms $A_{ij} \cong A'_{ij}$. And so on, for all n. There are two natural arrows

$$F(\mathfrak{U}_1) \rightrightarrows F(\mathfrak{U}_2)$$

The first arrow sends an element A of $F(\mathfrak{U}_1)$ to an element of $F(\mathfrak{U}_2)$ that assigns to each overlap $U_i \cap U_j$ the restriction $A_i|_{U_i \cap U_j}$. The second gives the element that assigns $A_j|_{U_i \cap U_j}$. Similarly, there are n natural arrows

$$F(\delta^n_i) : F(\mathfrak{U}_{n-1}) \to F(\mathfrak{U}_n)$$

for all n, assigning to each n-fold overlap the restriction of the configuration on the $(i+1)$th region.[4]

A collection of configurations covering X is an element A of $F(\mathfrak{U}_1)$. Intuitively, a *compatible* collection of configurations is one where A_U and A_V are the same when both are restricted to $U \cap V$, and A_U, A_V, and A_W are the same when all three are restricted to $U \cap V \cap W$, and so on. Rendered in categorial language, the set of collections of configurations satisfying this condition is the limit

$$\lim\Big(F(\mathfrak{U}_1) \rightrightarrows F(\mathfrak{U}_2) \substack{\rightarrow\\\rightarrow\\\rightarrow} F(\mathfrak{U}_3) \substack{\rightarrow\\\rightarrow\\\rightarrow\\\rightarrow} \cdots\Big)$$

Where these arrows are the arrows $F(\delta^n_i)$ above. A theory is separable when the state of X supervenes on the state of a cover for any cover. So we can further formalize separability as the following condition.[5]

Definition 2. A *sheaf* on a topological space X is a presheaf F on X such that for any cover $\mathfrak{U}$ of X, the

[4]There are also codegeneracy maps in the other direction. In brief, $F(\mathfrak{U}_\bullet)$ is the cosimplicial object obtained by composing F with the Čech nerve of the cover $\mathfrak{U}$; see Hollander (2008, §1.3).

[5]This is an admittedly nonstandard definition of the sheaf condition, chosen to make later generalization obvious. For the usual definition, see Mac Lane and Moerdijk (1992, 66). The definition there is equivalent to Def. 2.

restriction map

$$F(X) \to \lim\Big(F(\mathfrak{U}_1) \rightrightarrows F(\mathfrak{U}_2) \substack{\rightarrow\\\rightarrow\\\rightarrow} F(\mathfrak{U}_3) \substack{\rightarrow\\\rightarrow\\\rightarrow\\\rightarrow} \cdots\Big)$$

is an isomorphism.

This definition is rather more complicated than required for the application immediately at hand. Standard computations give the following simplification.[6]

Definition 3. For any topological space X with cover $\mathfrak{U}$ and presheaf F on X, let the *set of 0-descent data* $\mathrm{Desc}(\mathfrak{U}, F)$ be the subset of $F(\mathfrak{U}_1)$ consisting of elements A such that for all U_i and U_j in $\mathfrak{U}$, $A_i|_{U_i \cap U_j} = A_j|_{U_i \cap U_j}$.

Proposition 1. For any topological space X with cover $\mathfrak{U}$ and presheaf F on X, the usual construction of limits in Set gives

$$\mathrm{Desc}(\mathfrak{U}, F) \cong \lim\Big(F(\mathfrak{U}_1) \rightrightarrows F(\mathfrak{U}_2) \substack{\rightarrow\\\rightarrow\\\rightarrow} F(\mathfrak{U}_3) \substack{\rightarrow\\\rightarrow\\\rightarrow\\\rightarrow} \cdots\Big)$$

So the sheaf condition simplifies: $F(X)$ is a sheaf if $F(X) \cong \mathrm{Desc}(\mathfrak{U}, F)$ for all covers $\mathfrak{U}$ of X. In other words, F is a sheaf just in case a configuration of X is the same as a collection of configurations on subregions of X that agree on the overlaps, for any way of carving X up into subregions.

The sheaf condition further sharpens the definition of separability from the previous section. Essentially it formalizes what it means for the state of X to supervene on the states of its subregions. It also reproduces the arguments of Belot (1998), Healey (2007), and Myrvold (2011) against the separability of the holonomy interpretation. But it is not idle formalization. As we will see below, the sheaf condition is not the only way to formalize separability. This ambiguity has led to conflicting claims regarding the separability of fiber bundle interpretations of Yang–Mills theory: Healey (2007) takes them to be separable, while Wallace (2014) argues that they are not. This conflict can be resolved if we disambiguate the different analyses of separability. However, in doing so we uncover a similar ambiguity in the terms "fiber bundle formulation" and "holonomy formulation".

3 Two failures

Again, in less formal terms: a sheaf is an assignment of configuration spaces to regions such that a configuration of some region X is the same as a compatible collection of configurations of regions composing X. The state of the region X thus supervenes on the state of its subregions, because two different configurations in X correspond to two different compatible collections of configurations of regions composing X. This matches semiformal descriptions of separability. Furthermore, the sheaf condition delivers the same verdict on the

[6] "The usual construction of limits in Set" is the construction in terms of products and equalizers; see, e.g., Mac Lane (1998, 109).

Aharonov–Bohm effect regarding holonomy formulations of Yang–Mills theory: they are nonseparable. But for all that, I think the sheaf condition is not the right analysis of separability. It does not correctly treat gauge structure.

The Aharonov–Bohm setup demonstrates the failure of the sheaf condition for Healey's (2007) holonomy formulation, reflecting the standard arguments for its nonseparability.[7] Consider the punctured plane $\mathbb{R}^2 \setminus \{0\}$, and let

$$\left\| \mathsf{Hol}^{U(1)}(-) \right\| : \mathcal{O}(\mathbb{R}^2 \setminus \{0\})^{\mathrm{op}} \to \mathsf{Set}$$

be the presheaf that assigns to each region the set of $U(1)$-holonomy maps in that region. For the cover

$$U_\pm = \mathbb{R}^2 \setminus \{(x, 0) : \pm x \leq 0\}$$

the sheaf condition requires that a holonomy map on $\mathbb{R}^2 \setminus \{0\}$ be equivalent to the pair of holonomy maps on $U_\pm$ obtained by restriction. But it is not. For any value of the solenoid current, every holonomy map in the simply-connected regions $U_\pm$ is the trivial map. But different values of the solenoid current correspond to different holonomy maps. Thus the natural function

$$\left\| \mathsf{Hol}^{U(1)}(\mathbb{R}^2 \setminus \{0\}) \right\| \to \mathsf{Desc}\Big(\{U_\pm\}, \left\| \mathsf{Hol}^{U(1)}(-) \right\|\Big)$$

is not injective, hence not an isomorphism. So the sheaf condition performs as expected in this case, and for the expected reasons.

The sheaf analysis gives a similar verdict on fiber bundle formulations. This is somewhat surprising, given the standard story that contrasts separable fiber bundle formulations with non-separable holonomy interpretations. Nevertheless, this argument has precedent in Wallace (2014, §4). He argues that we have nonseparability in the setup above when we have potentials A and A' on $\mathbb{R}^2 \setminus \{0\}$ such that $A|_{U_\pm}$ is gauge-equivalent to $A'|_{U_\pm}$ but A is not gauge-equivalent to A'. In our terms, we are considering the presheaf

$$\left\| \Omega^1(-; \mathfrak{u}(1)) /\!/ U(1) \right\| : \mathcal{O}(\mathbb{R}^2 \setminus \{0\})^{\mathrm{op}} \to \mathsf{Set}$$

that assigns to each region the set of gauge-equivalence classes of vector potentials. Wallace's criterion judges this presheaf nonseparable if there are different equivalence classes in $\left\| \Omega^1(\mathbb{R}^2 \setminus \{0\}; \mathfrak{u}(1)) /\!/ U(1) \right\|$ that are

[7] Most arguments against separability show that the sheaf condition (or an informal version of it) fail for the configuration space of interest. An exception is the argument given by Myrvold (2011), who argues that the algebra of observables for a region is not determined in the natural way by the algebras of observables for its subregions. His argument is categorically dual to the argument in terms of states: he shows that the pre-cosheaf assigning algebras of observables to spacetime regions is not a cosheaf. Since observables are, broadly speaking, functions on state space, Myrvold's argument is roughly Isbell-dual to the argument given in the main text (cf. Lawvere, 2005, 17).

sent to the same elements of $\left\|\Omega^1(U_\pm;\mathfrak{u}(1)) /\!\!/ U(1)\right\|$. In other words, nonseparability occurs when the natural map

$$\left\|\Omega^1(\mathbb{R}^2 \setminus \{0\};\mathfrak{u}(1)) /\!\!/ U(1)\right\| \to \mathrm{Desc}\Big(\{U_\pm\}, \left\|\Omega^1(-;\mathfrak{u}(1)) /\!\!/ U(1)\right\|\Big)$$

fails to be injective. And the Aharonov–Bohm setup provides just such a case. No matter the setting of the solenoid current, the vector potential in $U_\pm$ will be gauge-equivalent to the trivial potential. But the potential on $\mathbb{R}^2 \setminus \{0\}$ depends on the current in the solenoid. Since these different configurations on $\mathbb{R}^2 \setminus \{0\}$ are sent to the same compatible collection of potential equivalence-classes on $U_\pm$, the sheaf condition fails.

The sheaf condition judges both fiber bundle formulations and holonomy formulations nonseparable. But there's something fishy about the failure of separability in the case of fiber bundle formulations. Holonomy formulations straightforwardly fail to be separable: the restriction map is not invertible. But showing that fiber bundle formulations are nonseparable requires taking gauge-equivalence classes, which is not obviously the right thing to do. Indeed, I expect that the intuitive separability of the Aharonov–Bohm setup in the fiber bundle formulation comes from thinking about the vector potential as an element of the presheaf

$$\Omega^1(-;\mathfrak{u}(1)) : \mathcal{O}(X)^{\mathrm{op}} \to \mathsf{Set}$$

that assigns to each region the set of vector potentials (i.e., $\mathfrak{u}(1)$-valued 1-forms) in that region, ignoring the gauge structure. And this presheaf *is* a sheaf. So if we took $\Omega^1(-;\mathfrak{u}(1))$ to be the configuration space for electromagnetism then by the sheaf-theoretic analysis electromagnetism would be separable after all. Again using $\|-\|$ to denote taking the equivalence classes, this means there is an isomorphism

$$\left\|\Omega^1(\mathbb{R}^2 \setminus \{0\};\mathfrak{u}(1)) /\!\!/ U(1)\right\| \to \left\|\mathrm{Desc}\Big(\{U_\pm\}, \Omega^1(-;\mathfrak{u}(1)) /\!\!/ U(1)\Big)\right\|$$

Separability facts thus depend delicately on just when we form equivalence classes of gauge potentials.

This delicate dependence points to a deeper problem with the sheaf condition. The sheaf condition tells us when presheaves are separable. But presheaves assign mere sets to regions, and configuration spaces generally have more structure than mere sets. In particular, a fiber bundle formulation of Yang–Mills theory is most naturally interpreted as assigning a collection of geometric objects *along with* a notion of isomorphism between them. Recovering a set from this data in order to define a presheaf requires either forgetting about the isomorphisms or taking isomorphism classes. But neither of these options is ideal; one adds spurious distinctions and one removes structure from the theory. In the case of Yang–Mills theories, taking gauge-equivalence classes eliminates precisely the separability structure of the theory.

This particular, systematic misbehavior of limits is ubiquitous in mathematics. In some cases it can be rectified by moving to an equivalent category in which they are better behaved, but sometimes this is impossible. The root of the problem is that there are multiple interacting notions of sameness. In our case of vector potentials, for example, we have the usual mathematical notion of equality of vector fields, according to which two potentials are physically the same if they assign the same $\mathfrak{g}$-valued covector to each point of spacetime. But we also say that two potentials are the same if they are related by a gauge equivalence. Limits require that we choose one notion and stick with it, leading to the problems above. Avoiding these problems prompts us to pick up the tools of abstract homotopy theory, which is the study of such interacting sameness structures (Riehl, 2014). In what follows we rectify the analysis of separability as the sheaf condition by including gauge structure. This amounts to incorporating the homotopy theory of groupoids.

4 Groupoids

If we are forced to consider assignments of sets to regions of spacetime, then we are faced with the problems of the previous section. But we're not forced to do so; we could just as well consider the assignment of groupoids to regions of spacetime instead.

Definition 4. A *groupoid* is a category in which all arrows are isomorphisms.

Indeed, much of the philosophical literature on the interpretation of Yang–Mills theories already involves groupoids. For example, Rosenstock and Weatherall (2015) argue that fiber bundle and holonomy interpretations have the same amount of structure, in the sense that a particular groupoid of fiber bundles is equivalent to a particular groupoid of holonomy maps. Weatherall (2015), too, uses groupoids to analyze relative amounts of structure in gauge theories. More generally, Rosenstock et al. (2015) and Halvorson (2012, 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. This section introduces some tools from the theory of groupoids, along with the groupoids of interest for Yang–Mills theory.

First, we generalize presheaves of sets to presheaves of groupoids. This requires a definition of a category of groupoids, which will be the codomain of a presheaf of groupoids. We define this category Grpd as follows. An object of Grpd is a groupoid and a morphism between groupoids is a functor. The category Grpd is naturally equipped with further structure. There is a nontrivial notion of 2-morphisms, or morphisms between morphisms, given by natural transformations. So the relevant notion of sameness for maps of groupoids is natural isomorphism. Similarly, the relevant notion of sameness for groupoids is not isomorphism, but the less restrictive equivalence of categories, which we denote with $\simeq$. In what follows we will never use isomorphism to compare groupoids, only equivalence, and (what amounts to the same thing) we will never use equality to compare morphisms of groupoids, only natural isomorphism. With this in mind, we generalize presheaves to presheaves of groupoids.[8]

Definition 5. A *presheaf of groupoids* on X is a map $F : \mathcal{O}(X)^{\mathrm{op}} \to \mathsf{Grpd}$.

In the next section we give an analysis of separability for presheaves of groupoids that generalizes the sheaf condition. But first we review some useful facts about groupoids.

Sets are a special kind of groupoid. That is, there is a full inclusion functor

$$\iota : \mathsf{Set} \hookrightarrow \mathsf{Grpd}$$

that sends a set X to a groupoid $\iota(X)$ whose objects are the elements of X and whose arrows are just the identity arrows $\mathrm{id}_x : x \to x$ for each $x \in X$. As such, any presheaf (of sets) F is naturally a presheaf of groupoids $\iota \circ F$. In what follows we simply assume that everything is a groupoid, silently inserting an ι where needed. This inclusion has a reflector, called the truncation functor.[9]

Definition 6. For any groupoid X, its *truncation* $\|X\|$ is the set of isomorphism classes of X, considered as a groupoid.

There is a natural map $X \to \|X\|$, which sends each element x of X to the isomorphism class of x in $\|X\|$. If X is a set being considered as a groupoid, then $X \simeq \|X\|$. So if X is equivalent to some set (considered as a groupoid), then X is equivalent to $\|X\|$. If Y is a groupoid such that Y is a equivalent to its truncation, and $\mathsf{Grpd}(X, Y)$ is the groupoid of maps from X to Y, then we have an equivalence

$$\mathsf{Grpd}(\|X\|, Y) \simeq \mathsf{Grpd}(X, Y)$$

given by composition with the natural map $X \to \|X\|$. This means that every function on a groupoid that takes values in some set—any real-valued function on configuration space, say—is equivalently a function on the set of isomorphism classes of that set. However, when we consider maps into general groupoids, as separability will require, a groupoid and its truncation may importantly differ.

The truncation functor forgets information. For example, let G be any group and consider the groupoid with a unique element and one arrow for each element of G. Call this groupoid BG. For any G, $\|BG\|$ is the

[8] If we equip Grpd with the further information about 2-morphisms, then we treat it as a 2-category, and a presheaf of groupoids is a pseudofunctor. If we instead equip it with the equivalences of categories as weak equivalences then we make it into category with weak equivalences, and this induces a notion of weak equivalence for the functor category $[\mathcal{O}(X)^{\mathrm{op}}, \mathsf{Grpd}]$, and we use this notion of equivalence to compare presheaves of groupoids. These are really two expressions of the same construction, and the former can be recovered from that latter by localizing at weak equivalences. These details will not be terribly important for our purposes. See Lack (2010) for the 2-categorical approach, and Hollander (2008, Th. 2.1 & Prop. 4.1) for the homotopical point of view.

[9] The truncation considered here is the 0th rung of a ladder of truncation functors (Lurie, 2009, Prop. 5.5.6.18). Elsewhere (e.g., in UFP, 2013) this functor is denoted $\|-\|_0$, and the unadorned $\|-\|$ is reserved for the (-1)-truncation. Since we will be concerned solely with 0-truncation in what follows, we will use $\|-\|$ for the 0-truncation functor.

groupoid **1**, the singleton set considered as a groupoid. So the functor $BG \to \|BG\|$ isn't faithful. In the terminology of the Baez–Dolan–Bartel classification (Baez and Shulman, 2010), it forgets "stuff". We can think of a groupoid as composed of cardinality facts (the number of isomorphism classes) and sameness facts (all the isomorphisms). Applying the truncation functor forgets exactly the sameness facts. More precisely, any groupoid X is equivalent (non-canonically) to one of the form

$$X \simeq \coprod_{[x]\in\|X\|} B\mathrm{Aut}(x)$$

where $\mathrm{Aut}(x)$ is the group of automorphisms of x. So if we are given the set of isomorphism classes of some groupoid and the automorphism group of each isomorphism class, then we can reconstitute the original groupoid by taking the coproduct. In the other direction, truncating any groupoid gives

$$\left\| \coprod_{[x]\in\|X\|} B\mathrm{Aut}(x) \right\| \simeq \coprod_{[x]\in\|X\|} \|B\mathrm{Aut}(x)\| \simeq \coprod_{[x]\in\|X\|} \mathbf{1}$$

The cardinality facts, which index the coproduct, are left untouched, while the sameness facts are all made trivial.

We now define some groupoids to represent fiber bundle formulations of Yang–Mills theory. In simple cases, it suffices to consider the groupoid whose elements are vector potentials and whose isomorphisms are gauge equivalences. This gives the following groupoid.[10]

Definition 7. Let X be a smooth manifold and G a Lie group with Lie algebra $\mathfrak{g}$. Define the *groupoid of $\mathfrak{g}$-valued 1-forms on X* to be the groupoid $\Omega^1(X;\mathfrak{g})/\!/G$ where

- an object is a $\mathfrak{g}$-valued 1-form on X
- an arrow $A \to A'$ is a function $g : X \to G$ such that

$$A' = g^{-1}Ag + g^{-1}\,dg$$

The groupoid of $\mathfrak{g}$-valued 1-forms on X is an adequate description of the Yang–Mills configuration space only if we restrict attention to the sector with no magnetic monopoles or instantons. To treat these, we must instead work with principal connections on principal G-bundles. We will use the following definition.[11]

[10] We restrict our attention in what follows to matrix Lie groups, so that gauge equivalences can be written as the familiar $g^{-1}Ag + g^{-1}\,dg$. This isn't necessary; everything is still true for a general Lie group if we instead write gauge equivalences as $\mathrm{Ad}_{g^{-1}}A + g^*\theta$ for Ad the adjoint action of G on $\mathfrak{g}$ and θ the Maurer–Cartan form.

[11] For a textbook treatment taking this definition as primary, see Nicolaescu (1999, Def. 8.1.2). Most any textbook covering principal connections notes that they can be described this way but opts for a more intrinsic definition—e.g., Kobayashi and Nomizu (1996, 52 and 66) and Kolář et al. (1993, 87 and 102).

Definition 8. For a smooth manifold X and Lie group G with Lie algebra $\mathfrak{g}$, define the *groupoid of principal G-bundles with connection on X* to be the groupoid $G\mathsf{Bun}^{\nabla}(X)$ where

- an object is a triple $(\mathfrak{U}, A, g)$ of an open cover $\mathfrak{U}$ of X, a collection of $\mathfrak{g}$-valued 1-forms $A \in \Omega^1(\mathfrak{U}_1; \mathfrak{g})$, and for every $U_i, U_j \in \mathfrak{U}$ a function $g_{ij} : U_i \cap U_j \to G$ such that

$$A_j = g_{ij}^{-1} A_i g_{ij} + g_{ij}^{-1}\, dg_{ij} \qquad \text{and} \qquad g_{ij} g_{jk} = g_{ik}$$

  Here A_i denotes the $\mathfrak{g}$-valued 1-form assigned to U_i by A, and we omit the restriction maps.

- an arrow $h : (\mathfrak{U}, A, g) \to (\mathfrak{U}', A', g')$ is for every $U_i \in \mathfrak{U}$ and $U'_j \in \mathfrak{U}'$ a function $h_{ij} : U_i \cap U'_j \to G$ such that

$$A'_j = h_{ij}^{-1} A_i h_{ij} + h_{ij}^{-1}\, dh_{ij} \qquad \text{and} \qquad h_{ij} g'_{j\ell} = g_{ik} h_{k\ell}$$

  where, again, the restriction maps are omitted.

The amount of data in this representation of principal G-bundles with connection is unwieldy. But intuitively it is straightforward: an object of $G\mathsf{Bun}^{\nabla}(X)$ is a covering of X by patches on each of which there is a $\mathfrak{g}$-valued 1-form, and such that on each overlap of patches the 1-forms agree. The only wrinkle is that by "agree" we here mean that there is a specified gauge transformation from one to the other.

In what follows we will focus on the presheaf of groupoids $G\mathsf{Bun}^{\nabla}$, spending no time with the details of any holonomy representation. This is no loss, because we can appeal to Rosenstock and Weatherall's (2015) main result to draw parallel conclusions about holonomy representations. Rosenstock and Weatherall show that a certain groupoid **PC** of principal bundles with connection is equivalent to a certain groupoid **Hol** of holonomy models. For any smooth manifold X and Lie group G, $G\mathsf{Bun}^{\nabla}(X)$ is equivalent to a certain subgroupoid of **PC**, so by their result there is a corresponding subgroupoid $\mathsf{Hol}^G(X)$ of **Hol** such that

$$G\mathsf{Bun}^{\nabla}(X) \simeq \mathsf{Hol}^G(X)$$

Because we have decided to only compare groupoids with equivalence, $G\mathsf{Bun}^{\nabla}(X)$ and $\mathsf{Hol}^G(X)$ are the same for our purposes. In particular, one satisfies the stack condition if and only if the other does, and truncation induces an isomorphism

$$\left\| G\mathsf{Bun}^{\nabla}(X) \right\| \cong \left\| \mathsf{Hol}^G(X) \right\|$$

So any claim made about $G\mathsf{Bun}^{\nabla}$ or its truncation applies just as well to the corresponding holonomy representation.[12]

[12]More precisely, **PC** and **Hol** are both naturally fibered over the product groupoid $\mathsf{Man} \times \mathsf{LieGrp}$ of smooth manifolds and Lie

The previous paragraph resolves one of the tensions discussed in the opening. The standard story has it that fiber bundles involve more structure than holonomy models, but Rosenstock and Weatherall show that there is a precise sense in which a particular fiber 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, $G\mathsf{Bun}^{\nabla}(X)$ and $\mathsf{Hol}^{G}(X)$ have the same amount of structure, and their truncations have the same amount of structure, but the $G\mathsf{Bun}^{\nabla}(X)$ and $\mathsf{Hol}^{G}(X)$ have more structure than their truncations. In particular, the truncation map forgets stuff. In the standard story, "the" fiber bundle representation is usually taken to be $G\mathsf{Bun}^{\nabla}(X)$, and "the" holonomy representation is usually taken to be $\left\|\mathsf{Hol}^{G}(X)\right\|$. So, properly interpreted, this story is right. It's just not the whole story. Worse, this terminology has not be consistently maintained. Wu and Yang (1975), for example, regularly slide back and forth between $\mathsf{Hol}^{G}(X)$ and its truncation, treating the differences as a matter of convenience rather than substance. And the details of their argument rely on the structure of $\mathsf{Hol}^{G}(X)$ that is lost in truncation. 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 truncation.

The second tension concerned locality and the separability of the fiber bundle interpretation. If we reflect on the gloss of $G\mathsf{Bun}^{\nabla}(X)$ given above, it is intuitively separable. Indeed, we might think of $G\mathsf{Bun}^{\nabla}(X)$ as a localization of $\Omega^1(X;\mathfrak{g})/\!/G$: a configuration in $G\mathsf{Bun}^{\nabla}(X)$ is just a compatible collection of configurations that look like they come from $\Omega^1(-;\mathfrak{g})/\!/G$. To make this precise, we need a definition of separability that applies to presheaves of groupoids, not just presheaves of sets. In the next section we give this definition and show that $G\mathsf{Bun}^{\nabla}$ satisfies it. Since $G\mathsf{Bun}^{\nabla}$ is equivalent to a presheaf of groupoids assigning holonomy models to regions, there is a separable holonomy formulation, too. However, composing either of these presheaves with the truncation functor gives a nonseparable theory. This demonstrates the sense in which the question of separability is orthogonal to the choice between fiber bundle and holonomy models.

5 Stacks

The problem with the sheaf condition lies in its use of a limit, because limits are not invariant under equivalence of groupoids. For a toy example, consider the groupoid $\mathbf{1}$, which has one object $\star$ and only the identity arrow, and the groupoid I, which has two objects 0 and 1 whose only nontrivial arrows are an arrow $s : 0 \to 1$ and

groups, so by the Grothendieck construction (Johnstone, 2002, Th. 1.3.6) their equivalence corresponds to a natural isomorphism of pseudofunctors $\mathsf{PC}, \mathsf{Hol} : \mathsf{Man}^{\mathrm{op}} \times \mathsf{LieGrp}^{\mathrm{op}} \to \mathsf{Grpd}$. Considering $\mathcal{O}(X)$ as a subcategory of Man and holding the Lie group fixed, this restricts to an equivalence of presheaves $\mathsf{PC}^{G}, \mathsf{Hol}^{G} : \mathcal{O}(X)^{\mathrm{op}} \to \mathsf{Grpd}$. The presheaf $G\mathsf{Bun}^{\nabla}$ is a third presheaf of groupoids on X, equivalent to both of these.

its inverse. There are two arrows $i_0, i_1 : \mathbf{1} \to I$, which send the unique object of $\mathbf{1}$ to 0 and 1, respectively. These arrows are naturally equivalent, and both are equivalences of categories. But the limit of the diagram

$$\mathbf{1} \xrightarrow{i_0} I \xleftarrow{i_1} \mathbf{1}$$

is the empty groupoid, while the limit of the diagram

$$\mathbf{1} \xrightarrow{\mathsf{id}_\mathbf{1}} \mathbf{1} \xleftarrow{\mathsf{id}_\mathbf{1}} \mathbf{1}$$

is the groupoid $\mathbf{1}$. This should not be; these diagrams are the same diagram, according to the relevant standard of sameness. I is equivalent to $\mathbf{1}$, so the second diagram is obtained from the first by replacing I with a different representative of the same groupoid. This is a much simplified version of the problem we encountered with limits of Yang–Mills configuration spaces: we only get the right limits if we choose representatives for our groupoids in which isomorphism and identity coincide. But this is a feature of the set-theoretic representation, not a feature of the groupoids.

We're forced to fiddle about with different representations in this way because limits can't see any structure of a groupoid beyond the underlying set structure. By moving to an equivalent groupoid with a minimal underlying set before taking limits, we are encoding the true sameness structure in the underlying set, where the limit can access it.[13] It would be better to correct the limit operation to take the groupoid structure into account. This corrected construction goes by the name of the homotopy limit. Plugging it into the definition of the sheaf condition gives the stack condition.[14]

Definition 9 (Hollander, 2008, Def. 1.3)**.** For any topological space X, a presheaf of groupoids $F : \mathcal{O}(X)^{\mathrm{op}} \to \mathsf{Grpd}$ is a *stack* if for any open cover $\mathfrak{U}$ of X,

$$F(X) \simeq \mathrm{holim}\Big(F(\mathfrak{U}_1) \rightrightarrows F(\mathfrak{U}_2) \substack{\rightarrow\\\rightarrow\\\rightarrow} F(\mathfrak{U}_3) \substack{\rightarrow\\\rightarrow\\\rightarrow\\\rightarrow} \cdots\Big)$$

Actually giving a definition of the homotopy limit requires machinery far too heavy to wheel in here.[15] We will make do with an intuitive picture and a formula for our case of interest. For a motivating example,

[13] More precisely, we are applying a fibrant replacement functor before taking the limit, so as to make the ordinary limit coincide with the homotopy limit.

[14] Stacks, or 2-sheaves, or sheaves of groupoids, admit of many different approaches. Vistoli (2005) gives a careful introduction in terms of fibered categories. The point of view taken here comes from Hollander (2008), who also proves its equivalence with Vistoli's approach. The virtue of Hollander's point of view, for our purposes, is the way it exposes the role of standards of sameness.

[15] For an introduction to homotopy limits, see Dwyer and Spaliński (1995) and Riehl (2014). Classic references are Bousfield and Kan (1972) and Hirschhorn (2003).

take again the case of the Aharonov–Bohm effect, and consider the diagram

$$\Omega^1(U_+;\mathfrak{u}(1))/\!\!/U(1) \longrightarrow \Omega^1(U_+\cap U_-;\mathfrak{u}(1))/\!\!/U(1) \longleftarrow \Omega^1(U_-;\mathfrak{u}(1))/\!\!/U(1)$$

Call this diagram D. According to the sheaf condition, the Aharonov–Bohm setup is separable if the configuration space $\Omega^1(\mathbb{R}^2\setminus\{0\};\mathfrak{u}(1))/\!\!/U(1)$ is isomorphic to the limit of this diagram, $\lim D$. The limit $\lim D$ is defined to be a groupoid such that for any groupoid X, the groupoid of functions from X to $\lim D$ is isomorphic to the groupoid of cones from X to D,

$$\mathsf{Grpd}(X,\lim D)\cong\mathsf{Cone}(X,D)$$

A cone from X to D is a collection of arrows

$$\begin{array}{ccc} X & \longrightarrow & \Omega^1(U_-;\mathfrak{u}(1))/\!\!/U(1) \\ \downarrow & \searrow & \downarrow \\ \Omega^1(U_+;\mathfrak{u}(1))/\!\!/U(1) & \longrightarrow & \Omega^1(U_+\cap U_-;\mathfrak{u}(1))/\!\!/U(1) \end{array}$$

such that everything commutes. So an element of $\lim D$ is a pair of elements

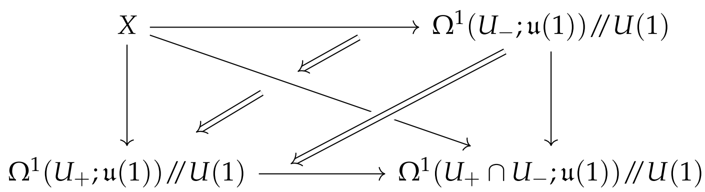
$$A_\pm\in\Omega^1(U_\pm;\mathfrak{u}(1)) \qquad \text{such that} \qquad A_+|_{U_+\cap U_-}=A_-|_{U_+\cap U_-}$$

So isomorphism of groupoids enters twice in the definition of the sheaf condition: once when we demand that the configuration space of some region is isomorphic to a certain limit, and once in the definition of the limit.

This is all wrong. The limit asks for two groupoids to be isomorphic, and it asks for any two paths in the diagram with the same start and end to be equal. But we shouldn't compare groupoids with isomorphism, and we shouldn't compare arrows between groupoids with equality. The right demand, which defines the homotopy limit of D, asks for a groupoid $\operatorname{holim} D$ such that there is an equivalence between maps from X to $\operatorname{holim} D$ and "homotopy coherent" cones from X to D,

$$\mathsf{Grpd}(X,\operatorname{holim} D)\simeq\mathsf{HoCone}(X,D)$$

In our case, a homotopy coherent cone is a diagram



in which everything commutes "up to homotopy". That is, it consists of three arrows out of X, along with natural isomorphisms—denoted by $\Rightarrow$ in the diagram—such that the composition of the shorter natural isomorphisms is equal to the long one.[16] It is not hard to show that $\Omega^1(\mathbb{R}^2 \setminus \{0\}; \mathfrak{u}(1)) /\!/ U(1)$ satisfies this universal property. That is, $\Omega^1(-; \mathfrak{u}(1)) /\!/ U(1)$ is separable on $\mathbb{R}^2 \setminus \{0\}$, according to the stack condition.[17]

For more general covers, we have a series of results that parallels the discussion of §2. For any cover $\mathfrak{U}$ and presheaf of groupoids F, we can give a formula for an object $\mathrm{Desc}(\mathfrak{U}, F)$ that computes the homotopy limit appearing in the stack condition. Showing that some presheaf of groupoids F is a stack then amounts to showing that it is equivalent to $\mathrm{Desc}(\mathfrak{U}, F)$ for any cover $\mathfrak{U}$.

Definition 10. Let X be a smooth manifold, with $\mathfrak{U}$ an open cover of X and F a presheaf of groupoids on X. Define $\mathrm{Desc}(\mathfrak{U}, F)$, the *groupoid of 1-descent data for F over* $\mathfrak{U}$, to be the groupoid where

- an object is a pair (A, g) where $A \in F(\mathfrak{U}_1)$ and g is an arrow in $F(\mathfrak{U}_2)$ such that for all $U_i, U_j, U_k \in \mathfrak{U}$ the diagram

$$\begin{array}{ccccc} & & A_j & & \\ & \overset{g_{ij}}{\nearrow} & & \overset{g_{jk}}{\searrow} & \\ A_i & & \xrightarrow[g_{ik}]{} & & A_k \end{array}$$

  commutes on $U_i \cap U_j \cap U_k$, and $g_{ii} = \mathrm{id}_{A_i}$.

- an arrow $h : (A, g) \to (A', g')$ is an arrow $A \to A'$ in $F(\mathfrak{U}_1)$ such that the square

$$\begin{array}{ccc} A_i & \xrightarrow{h_i} & A'_i \\ {\scriptstyle g_{ij}}\downarrow & & \downarrow{\scriptstyle g'_{ij}} \\ A_j & \xrightarrow[h_j]{} & A'_j \end{array}$$

  commutes on $U_i \cap U_j$ for all $U_i, U_j \in \mathfrak{U}$.

[16] For a real definition of homotopy-coherent diagrams, see Shulman (2009, Def. 10.2) or Riehl (2014, §7.7).

[17] Whether or not we should, we *can* ask whether $\Omega^1(-; \mathfrak{u}(1)) /\!/ U(1)$ satisfies the sheaf condition. It does on $\mathbb{R}^2 \setminus \{0\}$, but more generally it does not. It fails, for example, on $\mathbb{R}^3 \setminus \{0\}$.

Proposition 2 (Hollander, 2008, Cor. 2.11)**.** For any smooth manifold X, open cover $\mathfrak{U}$ of X, and presheaf of groupoids F on X,

$$\mathrm{Desc}(\mathfrak{U}, F) \simeq \mathrm{holim}\Big(F(\mathfrak{U}_1) \rightrightarrows F(\mathfrak{U}_2) \rightrightarrows F(\mathfrak{U}_3) \rightrightarrows \cdots\Big)$$

So to show that some presheaf of groupoids F is a stack, it suffices to show that $F(X) \simeq \mathrm{Desc}(\mathfrak{U}, F)$ for any cover $\mathfrak{U}$ of X. The intuitive reading of this condition is the same as the intuitive reading of the sheaf condition. F is a stack just in case a configuration of X is the same as a collection of configurations on subregions of X that agree on the overlaps, for any way of carving X up into subregions. All of the subtlety is found in what we mean by sameness.

The stack condition is a refinement of the sheaf condition, because they coincide when applied to sets. That is, suppose we have some presheaf $F : \mathcal{O}(X)^{\mathrm{op}} \to \mathsf{Set}$, and let $\iota : \mathsf{Set} \to \mathsf{Grpd}$ be the natural inclusion functor. Then

$$\mathrm{holim}\Big(\iota(F(\mathfrak{U}_1)) \rightrightarrows \iota(F(\mathfrak{U}_2)) \rightrightarrows \iota(F(\mathfrak{U}_3)) \rightrightarrows \cdots\Big) \simeq \lim\Big(F(\mathfrak{U}_1) \rightrightarrows F(\mathfrak{U}_2) \rightrightarrows F(\mathfrak{U}_3) \rightrightarrows \cdots\Big)$$

which is not hard to show by comparing the groupoids of descent data for each. So the argument from §3 continues to apply to the presheaves of groupoids

$$\Big\|\mathsf{Hol}^G(-)\Big\| : \mathcal{O}(X)^{\mathrm{op}} \to \mathsf{Grpd} \qquad \text{and} \qquad \Big\|G\mathsf{Bun}^\nabla(-)\Big\| : \mathcal{O}(X)^{\mathrm{op}} \to \mathsf{Grpd}$$

and these are still judged nonseparable. But this does not settle whether the presheaves of groupoids

$$\mathsf{Hol}^G : \mathcal{O}(X)^{\mathrm{op}} \to \mathsf{Grpd} \qquad \text{and} \qquad G\mathsf{Bun}^\nabla : \mathcal{O}(X)^{\mathrm{op}} \to \mathsf{Grpd}$$

are stacks, because applying the truncation eliminates groupoid structure that the homotopy limit makes use of.

$G\mathsf{Bun}^\nabla$ satisfies the stack condition. It is not difficult to show this directly, but it is rather tedious. We can take a shortcut by formalizing the intuitive description of $G\mathsf{Bun}^\nabla$ as a localized version of $\Omega^1(-;\mathfrak{g})/\!/G$. The localization process takes a presheaf of groupoids F and constructs a stack $\widetilde{F}$. There is an arrow $F \to \widetilde{F}$ that for each region U sends a configuration in $F(U)$ to a collection of local data determined by that configuration. This means that for any other stack F', an arrow $F \to F'$ should factor as $F \to \widetilde{F} \to F'$. After all, the stack F' is controlled by what happens locally, so if we first localize the data of a configuration with the arrow $F \to \widetilde{F}$ we shouldn't lose any of the relevant dependencies in the arrow $F \to F'$. This localization process is called stackification.

Proposition 3 (Laumon and Moret-Bailly, 2000, Lem. 3.2)**.** For any topological space X and presheaf of groupoids F on X, there is a stack $\widetilde{F}$, called the *stackification* of F, and a natural transformation $F \to \widetilde{F}$ that induces an equivalence

$$\mathsf{Sh}_X(\widetilde{F}, F') \simeq \mathsf{PSh}_X(F, F')$$

for all stacks F', where Sh_X and PSh_X are the categories of stacks and presheaves of groupoids on X, respectively.

Construction. For any $U \in \mathcal{O}(X)^{\mathrm{op}}$, let $\widetilde{F}(U)$ be the groupoid where

- an object is a triple $(\mathfrak{U}, A, g)$ of an open cover $\mathfrak{U}$ of U, a collection of objects $A \in F(\mathfrak{U}_1)$, and for every $U_i, U_j \in \mathfrak{U}$ an arrow $g_{ij} \in F(U_i \cap U_j)$ such that

$$\begin{array}{ccccc} & & A_j & & \\ & \overset{g_{ij}}{\nearrow} & & \overset{g_{jk}}{\searrow} & \\ A_i & & \xrightarrow[g_{ik}]{} & & A_k \end{array}$$

  commutes on $U_i \cap U_j \cap U_k$.

- an arrow $h : (\mathfrak{U}, A, g) \to (\mathfrak{U}', A', g')$ is for every $U_i \in \mathfrak{U}$ and $U'_j \in \mathfrak{U}'$ an arrow h_{ij} in $F(U_i \cap U'_j)$ such that the square

$$\begin{array}{ccc} A_i & \xrightarrow{h_{ij}} & A'_j \\ {\scriptstyle g_{ik}}\downarrow & & \downarrow{\scriptstyle g'_{j\ell}} \\ A_k & \xrightarrow[h_{k\ell}]{} & A'_\ell \end{array}$$

  commutes on $U_i \cap U'_j \cap U_k \cap U'_\ell$ for all $U_i, U_k \in \mathfrak{U}$ and $U'_j, U'_\ell \in \mathfrak{U}'$.

The arrow $F(U) \to \widetilde{F}(U)$ is given by

$$A \mapsto (\{U\}, A, \mathsf{id}_A)$$

♠

Proposition 4. $G\mathsf{Bun}^\nabla$ is the stackification of $\Omega^1(-;\mathfrak{g})/\!/G$, hence a stack.

The proof of Prop. 4 amounts to comparing the definitions of the terms involved. Because the property of being a stack respects equivalence, this also shows that Hol^G is a stack.

Based on these results, the argument of §3 can be turned around. One fiber bundle formulation of Yang–Mills theory has as its configuration presheaf $G\mathsf{Bun}^\nabla$, which is a presheaf of groupoids. The correct analysis of separability for a presheaf of groupoids is the stack condition, and $G\mathsf{Bun}^\nabla$ is a stack. So this fiber bundle formulation is separable.[18] So there are two kinds of fiber bundle formulations: ones like

[18] Just as the sheaf-theoretic argument against separability has Myrvold's dual argument in terms of observables, this stack-theoretic

$G\mathsf{Bun}^{\nabla}$—which incorporate the gauge structure of the theory and are separable—and the truncations of these—which eliminate the gauge structure and separability along with it. Holonomy formulations come in similar pairs of separable theories and truncated, nonseparable versions. So separability is orthogonal to the difference between fiber bundle and holonomy formulations. The association of holonomy representations with nonseparability rests on the coincidental focus in the literature on $\left\|\mathsf{Hol}^{G}(-)\right\|$ as the paradigmatic holonomy representation.

6 Conclusion

The primary geographic feature in the interpretive landscape of Yang–Mills theories has been a division between fiber bundle formulations and holonomy formulations. Other theoretical features—determinism, empirical underdetermination, locality, and more—have been mapped along this division. In particular, fiber bundle formulations have generally been treated as separable, while holonomy interpretations have not (Healey, 2007, §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 fiber bundles and holonomies. There are separable theories with configuration stacks

$$G\mathsf{Bun}^{\nabla} \simeq \mathsf{Hol}^{G}$$

and there are nonseparable theories with configuration presheaves

$$\left\|G\mathsf{Bun}^{\nabla}(-)\right\| \simeq \left\|\mathsf{Hol}^{G}(-)\right\|$$

So even after choosing between fiber 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 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 (though 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

argument for separability has a dual in terms of observables. For this, see Benini et al. (2015).

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, §1.1.1), too, has argued that non-perturbative 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 $G\mathsf{Bun}^{\nabla}(X)$ makes a difference to its standing with respect to other theories. It means that $G\mathsf{Bun}^{\nabla}(X)$ has the same amount of structure as $\mathsf{Hol}^{G}(X)$ and more than $\left\|\mathsf{Hol}^{G}(X)\right\|$, for example. But this groupoid structure also makes $G\mathsf{Bun}^{\nabla}$ separable, and this is just a fact about $G\mathsf{Bun}^{\nabla}$. So if we think that separability is a feature of theories—and it's hard to see what else it could be a feature of—then a theory must be more than its underlying set of models.

References

Arntzenius, F. (2012). *Space, Time, & Stuff*. Oxford University Press.

Baez, J. C. and May, J. P., editors (2010). *Towards Higher Categories*, volume 152 of *The IMA Volumes in Mathematics and its Applications*. Springer.

Baez, J. C. and Shulman, M. (2010). Lectures on *n*-categories and cohomology. In Baez and May (2010), pages 1–68.

Belot, G. (1998). Understanding electromagnetism. *The British Journal for the Philosophy of Science*, 49(4):531–555.

Benini, M., Schenkel, A., and Szabo, R. (2015). Homotopy colimits and global observables in abelian gauge theory. *Letters in Mathematical Physics*, 105(9):1193–1222.

Bousfield, A. K. and Kan, D. M. (1972). *Homotopy limits, completions, and localizations*, volume 304 of *Lecture Notes in Mathematics*. Springer.

Dwyer, W. G. and Spaliński, J. (1995). Homotopy theories and model categories. In James, I. M., editor, *Handbook of Algebraic Topology*, chapter 2, pages 73–126. Elsevier Science.

Einstein, A. (1935). Letter to Schrödinger, June 19. Cited in Howard (1985).

Einstein, A. (1948). Quantum mechanics and reality. *Dialectica*, 2:320–324. Cited in Howard (1985).

Halvorson, H. (2012). What scientific theories could not be. *Philosophy of Science*, 79(2):183–206.

Halvorson, H. (2015). Scientific theories. Forthcoming in *Oxford Handbooks Online*.

Healey, R. (2004). Gauge theories and holisms. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, 35(4):619–642.

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

Hirschhorn, P. S. (2003). *Model categories and their localizations*, volume 99 of *Mathematical Surveys and Monographs*. American Mathematical Society.

Hollander, S. (2008). A homotopy theory for stacks. *Israel Journal of Mathematics*, 163(1):93–124.

Howard, D. (1985). Einstein on locality and separability. *Studies in History and Philosophy of Science*, 16:171–201.

Johnstone, P. T. (2002). *Sketches of an Elephant: A Topos Theory Compendium*, volume 1. Clarendon Press.

Kobayashi, S. and Nomizu, K. (1996). *Foundations of Differential Geometry*. Wiley-Interscience.

Kolář, I., Michor, P. W., and Slovák, J. (1993). *Natural Operations in Differential Geometry*. Springer-Verlag.

Lack, S. (2010). A 2-categories companion. In Baez and May (2010), pages 105–192.

Laumon, G. and Moret-Bailly, L. (2000). *Champs algébriques*. Springer.

Lawvere, F. W. (2005). Taking categories seriously. *Reprints in Theory and Applications of Categories*, 8:1–24.

Leeds, S. (1999). Gauges: Aharonov, Bohm, Yang, Healey. *Philosophy of Science*, 66:606–627.

Lurie, J. (2009). *Higher Topos Theory*. Princeton University Press.

Lyre, H. (2004). Holism and structuralism in $U(1)$ gauge theory. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, 35(4):643–670.

Mac Lane, S. (1998). *Categories for the Working Mathematician*. Springer.

Mac Lane, S. and Moerdijk, I. (1992). *Sheaves in Geometry and Logic*. Springer.

Myrvold, W. C. (2011). Nonseparability, classical, and quantum. *The British Journal for the Philosophy of Science*, 62(2):417–432.

Nicolaescu, L. I. (1999). *Lectures on the Geometry of Manifolds*. World Scientific.

Redhead, M. (2001). The intelligibility of the universe. *Royal Institute of Philosophy Supplement*, 48:73–90.

Riehl, E. (2014). *Categorical Homotopy Theory*. Number 24 in New Mathematical Monographs. Cambridge University Press.

Rosenstock, S., Barrett, T. W., and Weatherall, J. O. (2015). On Einstein algebras and relativistic spacetimes. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, 52:309–316.

Rosenstock, S. and Weatherall, J. O. (2015). A categorical equivalence between generalized holonomy maps on a connected manifold and principal connections on bundles over that manifold. arXiv:1504.02401.

Schreiber, U. (2013). Differential cohomology in a cohesive ∞-topos. arXiv:1310.7930v1.

Shulman, M. (2009). Homotopy limits and colimits and enriched homotopy theory. arXiv:0610194.

The Univalent Foundations Program (2013). *Homotopy Type Theory: Univalent Foundations of Mathematics*. http://homotopytypetheory.org/book, Institute for Advanced Study.

Vistoli, A. (2005). Grothendieck topologies, fibered categories and descent theory. In Fantechi, B., Göttsche, L., Illusie, L., Kleiman, S. L., Nitsure, N., and Vistoli, A., editors, *Fundamental Algebraic Geometry: Grothendieck's FGA Explained*, volume 123 of *Mathematical Surveys and Monographs*. American Mathematical Society.

Wallace, D. (2014). Deflating the Aharonov–Bohm effect. arXiv:1407.5073.

Weatherall, J. O. (2015). Understanding gauge. Forthcoming in *Philosophy of Science*.

Wu, T. T. and Yang, C. N. (1975). Concept of nonintegrable phase factors and global formulation of gauge fields. *Physical Review D*, 12:3845–3857.