Gauge Invariance through Gauge Fixing

David Wallace

April 29, 2024

Abstract

Phenomena in gauge theory are often described in the physics literature via a specific choice of gauge. In foundational and philosophical discussions this is often criticized as introducing gauge dependence, and contrasted against (often aspirational) "gauge-invariant" descriptions of the physics. I argue, largely in the context of scalar electrodynamics, that this is misguided, and that descriptions of a physical process within a specific gauge are in fact gauge-invariant descriptions. However, most of them are *non-local* descriptions of that physics, and I suggest that this ought to be the real objection to such descriptions. I explore the unitary gauge as the exception to this nonlocality and consider its strengths and limitations, as well as (more briefly) its extension beyond scalar electrodynamics.

1 Introduction: who's afraid of gauge fixing?

Consider the theory of a complex scalar field interacting electromagnetically ('scalar electrodynamics', as particle theorists call it). In this theory, the fields are represented by ordered pairs (ψ, A), where ψ is a complex scalar field on Minkowski spacetime on $\mathcal{M}$ and A is a covariant vector field (i.e., one-form field). We normally also include some boundary condition: asymptotically at spatial infinity, A tends to zero and ψ tends to some constant value.[1] Theories like this are varyingly used to describe classical field theories, the theory of a quantum particle moving in a background classical electromagnetic field, and (normally through the mechanism of the path integral, and/or using the formalism of the quantum effective action) scalar electrodynamics, the theory of a complex quantum scalar field interacting with a quantum electromagnetic field. I write $\mathcal{F}$ for the space of all such fields satisfying the boundary conditions and whatever smoothness requirements a particular theory leads us to impose.

It is key to theories of this kind that their dynamics (however formulated) is *gauge-invariant*: invariant, that is, under transformations of the form

$$\psi(x) \to U(x)\psi(x), \quad A_\mu(x) \to A_\mu(x) - \frac{1}{e}U^{-1}(x)\partial_\mu U(x) \tag{1}$$

---

[1]In some applications of gauge theory — such as $SU(2)$ gauge theory in three spatial dimensions — it is possible to have a spatially varying boundary condition (see, e.g., (Cheng and Li 1984, section 15.3); I set this aside for simplicity.

where U is a smooth complex function on spacetime satisfying $|u(x)| = 1$, or equivalently (since Minkowski spacetime is simply connected),

$$\psi(x) \to \mathrm{e}^{ie\theta(x)}\psi(x), \quad A_\mu(x) \to A_\mu(x) - \partial_\mu\theta(x) \tag{2}$$

where θ is a smooth real function on spacetime. (In either case, e is a nonzero real number representing the strength of the electromagnetic coupling.) These transformations are called *local $U(1)$* transformations, since U can also be thought of as a function from spacetime to the group $U(1)$; the group of all such transformations is $\mathcal{G}$.

This framework readily generalizes to *non-Abelian gauge theory*, where now ψ is not just a complex field but takes values in some space $\mathcal{V}$, and where A is a one-form taking values in a Lie algebra of operators on $\mathcal{V}$; it also generalizes to theories where ψ is a spinor or vector field. In each case there is a generalization of the notion of a gauge group, with U taking values in a group larger than $U(1)$ and applying a somewhat more complicated transformation rule to A. In this general framework, if G is the group in which U takes values, the gauge group $\mathcal{G}$ is the group of all such U, i. e. all smooth maps from spacetime to G. The technical details are well known (see, e. g. , (Weinberg 1995, ch.15)) but will not be needed here. (It is also common in the philosophical literature — see, e.g., (Healey 2007; Weatherall 2016) — to use the more elegant, but more abstract, framework of fiber bundles, so that ψ becomes a section of a $\mathcal{V}$-bundle over spacetime, A becomes a connection on that bundle, and $\mathcal{G}$ becomes a group of automorphisms of that bundle but for my purposes this is unnecessary.)

Gauge symmetry in physics is normally taken as a sign of *representational redundancy*: the pair (ψ, A) overdescribes the physical goings on, so that distinct such pairs actually describe the same physics. To be more precise: let's say that a gauge transformation is *small* if its defining function U tends to the identity at infinity and if it can be smoothly deformed to the identity while preserving this boundary condition. (For $U(1)$ gauge theory on Minkowksi spacetime, this second condition is actually redundant, but for some gauge theories, and some background topologies, there are gauge transformations that equal the identity at the boundary but cannot be deformed to the identity everywhere while respecting the boundary condition.) And a transformation is *boundary-preserving* if it leaves the spatial boundary condition invariant. Any small gauge transformation is boundary-preserving, but the converse need not be true. Small and boundary-preserving transformations each form a group, which I write respectively as $\mathcal{G}_S$ and $\mathcal{G}_{BP}$.

Then the normal interpretation of gauge transformations in the physics literature is:

- Small gauge transformations simply encode representational redundancy: fields related by a small gauge transformation are the same field differently described. The group $\mathcal{G}_S \subset \mathcal{G}$ of small gauge transformations is not a *physical* symmetry group at all, but just a representation of that redundancy.

- Large gauge transformations, insofar as they are boundary-preserving (that is, insofar as they are defined at all as transformations on the space of fields) are physically-meaningful symmetry transformations in the same way as (say) the global $U(1)$ symmetries of the Klein-Gordon field, or the translation and rotation symmetries of Euclidean space. The quotient group $\mathcal{G}_{BP}/\mathcal{G}_S$ (whose action on a classical field is defined up to a small gauge transformation) is the physical (internal) symmetry group of the theory.

(I defend this physics orthodoxy *in extenso* in Wallace (2022c, 2022a, 2022b); see also (Belot 2018). Here I just assume it.)

At a somewhat abstract level, the true space of classical fields is not $\mathcal{F}$ but the quotient space $\mathcal{F}/\mathcal{G}_S$, whose elements are equivalence classes under the equivalence relation $(\psi, A) \sim (\psi', A')$, which holds between two field pairs iff they are related by a small gauge transformation. But this space is hard either to calculate with concretely, or to interpret physically. In mainstream physics, the normal route to address this problem is *gauge fixing*: we come up with some condition on field pairs (ψ, A) such as to pick out exactly one element of each equivalence class. It will be helpful to have some specific examples[2] (in each case from scalar electrodynamics):

Coulomb gauge: Choose a space/time separation on Minkowski spacetime, write $A = (V, \vec{\mathbf{A}})$, and impose the condition that $\vec{\nabla} \cdot \vec{\mathbf{A}} = 0$. To see that this does successfully define a gauge fixing, consider an arbitrary field $(\psi, V, \vec{\mathbf{A}})$ and look for the gauge transformation $(\psi, V, \vec{\mathbf{A}}) \rightarrow (\mathrm{e}^{ie\theta}\psi, V - \dot{\theta}, \vec{\mathbf{A}} - \vec{\nabla}\theta)$ that puts it into Coulomb gauge. We can easily see that θ solves

$$\nabla^2 \theta = \vec{\nabla} \cdot \vec{\mathbf{A}} \tag{3}$$

which has a guaranteed solution that is unique given the boundary conditions. So the gauge fixing condition picks out *at least* one element of each equivalence class (because some such θ can always be found) but also *at most* one element (because θ is unique). Note that Coulomb gauge is purely a condition on the vector-field part of the field pair and does not constrain the matter field.

Lorenz gauge: require $\partial_\mu A^\mu = 0$. By essentially the same argument as for Coulomb gauge, any field is related to a field in Lorenz gauge by a gauge transformation unique up to some function χ satisfying $\partial_\mu \partial^\mu \chi = 0$. Given some boundary conditions at infinity, this makes Lorenz gauge a valid gauge transformation provided that it is supplemented by some additional gauge-fixing term on the initial (or final) data (e.g. that it satisfies $\partial_i A^i = 0$ at the time). Lorenz gauge is again a condition purely on the vector field.

Unitary gauge: require that ψ is real and positive. Since we can write $\psi = \rho \mathrm{e}^{i\phi}$, it is immediate for nonzero ρ that the gauge transformation $\theta = \phi/e$

---

[2]For details, see any modern quantum field theory textbook, e. g. (Weinberg 1995; Peskin and Schroeder 1995).

uniquely transforms a field into unitary gauge. Unitary gauge, unlike our first two examples, is a condition purely on the scalar field, leaving the vector field unconstrained. (Unitary gauge has some subtleties, especially in the case where $\rho = 0$, which I consider in more detail later.)

R_ξ **gauge:** Writing $\psi(x) = (\rho_0 + v + ib)$ for some nonnegative real number ρ_0 (usually whatever the boundary condition is) we impose

$$\partial_\mu A^\mu = e\xi vb \tag{4}$$

for arbitrary ξ. R_ξ gauge is a hybrid of unitary and Lorenz gauge, and places a joint constraint on scalar and vector fields. (I leave it to the reader to demonstrate that it successfully fixes a gauge, given appropriate initial and boundary conditions.)

The normal assumption in physics is that gauge fixing is necessary to remove the descriptive redundancy, but that the specific choice of gauge fixing is a situation-dependent matter of convenience and pragmatics. The Coulomb gauge is excellently suited to reveal the degrees of freedom of the electromagnetic field, but is awkward calculationally because it breaks Lorentz invariance; Lorenz gauge is much more convenient in that respect. Unitary gauge is conceptually helpful in understanding the Higgs mechanism, but R_ξ gauge is better suited to addressing renormalization in Higgs calculations.[3] A choice of gauge, just like a choice of coordinate system, is a necessary preliminary to doing physics.

In the foundational literature, and in philosophy of physics, the mainstream physics approach has been met with suspicion. The general concern is that since only the gauge-invariant features of the field are physically meaningful, any gauge-fixed description mixes up genuinely gauge-invariant (and so, real) features of the physical problem with gauge-dependent (and so, illusory) features that just reflect our conventional choice of one gauge-fixing over another. The main foci of this literature are the Higgs mechanism and the Aharonov-Bohm (A-B) effect. In the former case, the definitive statement of the objection remains John Earman's ((Earman 2003, 189-190)):

> The popular presentations use the slogan that the vector field has acquired its mass by "eating" the Higgs field. But, as the authors of the standard treatises well know but rarely bother to warn the unwary reader, talk of *the* Higgs field has to be carefully qualified since by itself, the value of $[\psi]$ does not have gauge invariant significance. The popular slogan can be counterbalanced by the cautionary slogan that neither mass nor any other genuine attribute can be gained by eating descriptive fluff.... it is dereliction of duty for philosophers to repeat the physicists' slogans rather than asking what is the content of the reality that lies behind the veil of gauge.

In the latter, one of the early influential statements is Healey's (1997, p.22):

---

[3]See, e. g. , (Weinberg 1995, section 21.2) for a discussion of this point.

> [T]he A-B effect is local only if [the vector potential] is a physically real field, capable of acting on the electrons directly. But there is reason to doubt that the magnetic vector potential is a physically real field, since [it] is not gauge-invariant, unlike the magnetic field $\vec{\mathbf{B}}$ and the phase-shift...

The search for gauge-invariant accounts of the A-B effect is worth reviewing in more detail, as a reference point for later considerations. The usual argument goes: in classical electromagnetism the field strength

$$F_{\mu\nu} = \partial_\nu A_\mu - \partial_\mu A_\nu \tag{5}$$

(or its electric and magnetic components

$$\vec{\mathbf{E}} = \vec{\mathbf{A}} - \vec{\nabla}V; \quad \vec{\mathbf{B}} = \vec{\nabla} \times \vec{\mathbf{A}} \tag{6}$$

in any given inertial frame) are gauge-invariant and provide a complete description of the gauge-invariant features of the theory, but this ceases to be true when we consider the interaction of an electric field with a complex charged field, whether understood classically or through quantum particle mechanics or QFT. For consider a static magnetic field defined on a space coordinatized with Lorentz-frame coordinates (t, x, y, z) from which the z-axis has been removed from space at each instant in time (this 'solenoid spacetime' might be used to model an idealized solenoid). We take the electromagnetic field $A = (V, \vec{\mathbf{A}})$, in Coulomb gauge, to be

$$V(t, z, y, z) = 0; \quad \vec{\mathbf{A}}(x, y, z) = \frac{\Phi}{2\pi(x^2 + y^2)}(y, -x, 0) \tag{7}$$

We can readily calculate that $\vec{\mathbf{E}} = \vec{\mathbf{B}} = 0$, but the quantity

$$\Delta = \oint_C \vec{\mathbf{A}} \cdot \mathrm{d}\vec{\mathbf{x}} = \Phi \tag{8}$$

(where C is any path enclosing the missing z-axis) is equally readily demonstrated to be gauge-invariant.[4] And this quantity is physically significant: Δe is exactly the induced phase shift for interference between beams of charged particles that pass to either side of the solenoid. So we have gauge-inequivalent A-fields, but that inequivalence is invisible as long as we have only the field strengths. If we de-idealize the problem by restoring a physical solenoid along the z-axis, we need to thread a magnetic field of total flux Φ (strictly speaking, $\Phi + n/e$) through a narrow tube around the axis, but the magnetic field remains zero in the region through which the charged particles actually pass. So we seem to be left with a trilemma (explored in detail by Healey (2007)):

---

[4]More accurately, it is invariant under *small* gauge transformations; large gauge transformations can modify it by integer quantities.

1. We could take the field to be characterized by the field strengths alone — but then that characterization is inadequate for spacetimes of certain topology, and commits us to a form of action at a distance even in Minkowski spacetime.

2. We could take it to be characterized by the holonomies

$$H[C] = \oint_C A_\mu \cdot \mathrm{d}x^\mu \qquad (9)$$

   where now C is any closed path in spacetime — but now our characterization (as well as being extremely awkward calculationally) is non-separable, since the separate holonomies in regions A and B will underdetermine the holonomies in $A \cup B$.

3. Or we could just take it to be characterized by the field A — but now that characterization is gauge-dependent, and (so long as we hold on to the idea that gauge-related fields are just different representations of the same physical goings-on) our description obscures matters by mixing up genuine physics and descriptive fluff.

I pause to notice an oddity about this (widely used) approach to the A-B effect. It assumes *vacuum* electrodynamics, or else electrodynamics against a background of externally-imposed charge and current densities, so that we have to characterize only the electromagnetic field, and can neglect the charged matter field. But the equations of classical electrodynamics are

$$\partial_\nu F^{\mu\nu} = 4\pi J^\mu \qquad (10)$$

where the current J^μ is either stipulated as background or dynamically generated from some gauge-invariant phenomenological model, such as charged dust. And the (internal) dynamical symmetry group of this theory is not just local $U(1)$ symmetries, but the wider group of symmetries

$$A \to A + \chi \qquad (11)$$

where χ is any closed one-form field: that is, any one-form field satisfying $\partial_\mu \chi_\nu + \partial_\nu \chi_\mu$. (On a simply-connected manifold a one-form is closed if and only if it is exact (can be expressed as the derivative of a function) but this is not the case on multiply-connected manifolds, like our idealized-solenoid spacetime.) Holonomies are not invariant under these transformations; indeed, it is easy to show that the A-field on a region is fully characterized, up to transformations of form (11), by the field strength F on that same region, so that after all the field strength provides a local, gauge-invariant characterization. It is exactly because electromagnetism is invariant under this wider symmetry group that the Aharonov-Bohm effect requires quantum mechanics (or at any rate coupling to a complex field that also transforms under local $U(1)$). But in that context, we should be looking for a gauge-invariant description not just of A but of the

pair (ϕ, A) — and neither the field strength nor the holonomy does that. I expand on this observation in (Wallace 2014).

I return to the issue of characterizing (ψ, A) and not just A in section 3. For now I want to point out a more basic problem with the idea that we need to replace gauge-fixed accounts of electromagnetic physics with genuinely gauge-invariant ones. Consider again the gauge-fixed expression (7) for the solenoid spacetime's scalar and vector potential in the Coulomb gauge, and pick a specific t, x, y, z. The property that an electromagnetic potential has if, in the Coulomb gauge, it has value $(\Phi/2\pi(x^2 + y^2))(0, y, -x, z)$ at point (t, x, y, z) *is a gauge invariant property*.

How can this be? Well: a gauge-invariant property is one that holds of a field (ψ, A) iff it holds of any gauge-related field $(u_*(\psi, A))$. The statement that the A-field at (t, x, y, z) has such-and-such value, *simpliciter*, is obviously not gauge-invariant: any gauge transformation with nonzero derivative at (t, x, y, z) will change it. But two fields that are gauge related, are, by definition, exactly the same when expressed in a specific gauge.

We can go further. The statement "there is a gauge in which the A-field is given by (7)" is also gauge-invariant. It holds of a field A iff there is a gauge transformation that takes it to (7). But of course if A and A' are gauge-related then this holds of one of them iff it holds of the other.

(There is a close analogy to the equally-counter-intuitive fact that we can use coordinate systems to give coordinate-independent specifications of the distribution of matter. One way to say that a given curve in $\mathbb{R}^2$ is diffeomorphic to the circle, for instance, is to say that there is a smooth coordinatisation of $\mathbb{R}^2$ in which the curve satisfies $x^2 + y^2 = 1$. I expand on this feature of coordinate systems in (Wallace 2019).)

So: gauge-fixing is not an alternative to a gauge-invariant characterization of the fields of scalar electrodynamics; it is a way to provide a gauge-invariant characterization. And insofar as there is reason to find such characterizations unsatisfactory, that reason cannot be that they are gauge-variant.

Let us see what it might be.

2 A framework for gauge-invariant descriptions

It will be helpful to consider the problem in a more general and abstract framework. Let's say that a *gauge field theory* is a triple $(\mathcal{M}, \mathcal{F}, \mathcal{G}_S)$, where

- $\mathcal{M}$ is a manifold (possibly with additional structure);

- $\mathcal{F}$ is a space of fields on $\mathcal{M}$. I leave intentionally vague exactly what 'field' means here: certainly I intend it to cover at least the usual collections of scalar, vector, tensor, and spinor fields one finds in particle physics. But I require fields to satisfy two conditions:

  (i) *Fields are local:* if φ is a field on $\mathcal{M}$, it has a well-defined restriction to a field $\varphi|_X$ on any open subset $X \subset \mathcal{M}$, satisfying the consistency requirement that $(\varphi|_X)|_Y = \varphi|_Y$ for $Y \subset X \subset \mathcal{M}$.

(ii) *Fields are separable:* If $\{X_i\}$ is a collection of open subsets of $\mathcal{M}$ with $\cup_i X_i = X$, then for any two fields φ, φ', if $\varphi|_{X_i} = \varphi'|_{X_i}$ for each i then $\varphi|_X = \varphi'|_X$. (In other words, a field is specified in a region by specifying its values in elements of an arbitrarily small partition of the region.)

$\mathcal{F}_X$ is the space of fields restricted to X.

- $\mathcal{G}_S$ is a group of bijections of $\mathcal{F}$ — the (small) gauge transformations of the gauge field theory. Any element $g \in \mathcal{G}_S$ has a restriction $g|_X$ to any open subset $X \subset \mathcal{M}$, and this restriction commutes with field restriction: $(g|_X)_*(\varphi|_X) = (g_*\varphi)|_X$. Two restricted fields φ, φ' on X are gauge-equivalent iff there is some $g \in \mathcal{G}_S$ such that $g|_A(\varphi) = g|_A(\varphi')$; as a special case, two fields on the whole manifold are gauge-equivalent iff they are transformed into one another by some $g \in \mathcal{G}_S$.

Of course, this is a very attenuated notion of physical field, lacking any notion of dynamics, but it will suffice for our purposes. Scalar electrodynamics clearly fits this framework: the fields are pairs (φ, A) satisfying the boundary conditions, the small gauge transformations are the small local $U(1)$ transformations. So do non-Abelian gauge theories, and theories with spinor or tensor matter fields.

Given this setup, a *descriptor* d for the theory is a map $d : \mathcal{F} \to \mathcal{D}$, where $\mathcal{D}$ is any space we like. A descriptor might return a field's value at some spacetime point, or its average over some spacetime region, or its asymptotic behavior, or it might be a map taking values in $\{T, F\}$, returning T iff the field has some specified property.

A descriptor d is *gauge-invariant* if $d(\varphi') = d(\varphi)$ whenever φ' and φ are gauge-equivalent, i. e. whenever there is some $g \in \mathcal{G}_S$ such that $g_*\varphi = \varphi'$. Conversely, a descriptor is *complete* if whenever $d(\varphi) = d(\varphi')$, φ and φ' are gauge-equivalent. A complete gauge-invariant descriptor provides a non-redundant, gauge-invariant description of the field theory: if we know that a field is mapped to some $x \in \mathcal{D}$, we know which field it is up to gauge invariance.

In electrodynamics, each value $F(x)$ of the field strength at spacetime point x is a descriptor; so is the overall field F. Both are gauge-invariant; the latter is a complete descriptor for vacuum electrodynamics provided that the spacetime is simply connected. Each holonomy $\Delta(C)$ is a (gauge-invariant) descriptor; so is the holonomy map $\mathcal{C} \to \Delta(\mathcal{C})$; the latter is complete for vacuum electrodynamics whatever the topology of the spacetime. The vector field itself at a point, $A(x)$, is a descriptor; so is the matter field $\psi(x)$; so is the (trivial) descriptor that just assigns each field pair (ψ, A) to itself. The latter is complete; neither are gauge-invariant.

Gauge fixings are also easily described in this framework: a gauge-fixing f is a map from $\mathcal{F}$ to itself with the properties that (i) $f(\varphi)$ is gauge-equivalent to φ, and (ii) if φ and φ' are gauge-equivalent, $f(\varphi) = f(\varphi')$. (This is just to say that a gauge-fixing picks out a preferred member of each gauge equivalence class.

My point in the previous section is now simply that *gauge fixings are complete gauge-invariant descriptors*. This is easy to demonstrate, indeed almost immediate. Recall that a gauge fixing selects one element from each class of gauge-equivalent fields. So it can be represented by a map f from $\mathcal{F}$ to itself, such that $f(\varphi) = f(\varphi')$ iff φ and φ' are gauge-equivalent. But that is just the condition for f to be a complete gauge-invariant descriptor.

So what — if anything — is wrong with just saying that we extract a theory's gauge-invariant content just by choosing a gauge?

3   Locality

To answer this, let's specialize again: back to scalar electrodynamics, and specifically to Coulomb gauge. We have

$$\vec{\nabla} \times (\vec{\nabla} \times \vec{\mathbf{A}}) \equiv \vec{\nabla} \times \vec{\mathbf{B}}; \tag{12}$$

using the identity $\vec{\nabla} \times \vec{\nabla} \times \vec{\mathbf{A}} = \vec{\nabla}(\vec{\nabla} \cdot \vec{\mathbf{A}}) - \nabla^2 \vec{\mathbf{A}}$ and the Coulomb gauge condition, we get

$$\nabla^2 \vec{\mathbf{A}} = \vec{\nabla} \times \vec{\mathbf{B}}. \tag{13}$$

The right side of this equation is gauge-invariant, and (given boundary conditions) the equation can be solved explicitly to give the left side in terms of the right:

$$\vec{\mathbf{A}}(\vec{\mathbf{x}}, t) = \frac{1}{4\pi} \int \mathrm{d}x'^3 \, \frac{\vec{\nabla} \times \vec{\mathbf{B}}(\vec{\mathbf{x}'})}{|\vec{\mathbf{x}} - \vec{\mathbf{x}'}|} \tag{14}$$

On the one hand, this (combined with the equivalent result for V) is an explicit definition of my general observation that a gauge-fixed potential is uniquely determined by gauge-invariant features of the field, and so is itself gauge-invariant.

On the other hand, it is equally true that the gauge-fixed potential is *non-locally* dependent on manifestly local gauge-invariant features of the field like $\vec{\mathbf{B}}$. To determine $\vec{\mathbf{A}}(t, \vec{\mathbf{x}})$ in the Coulomb gauge we need to know the magnetic field arbitrarily far from $\vec{\mathbf{x}}$ (albeit the degree of dependence falls off with the inverse square of distance). $\vec{\mathbf{A}}(t, \vec{\mathbf{x}})$ may be a gauge-invariant quantity, but it is not in any physically meaningful sense localized at $(t, \vec{\mathbf{x}})$, despite its deceptive formulation as a vector field.

With this in mind, let's return to our abstract formalism and ask how to think about locality. We can define a *local descriptor d* as follows: it is specified (for a gauge field theory $(\mathcal{M}, \mathcal{F}, \mathcal{G}_S)$) by:

- A map $X \to D_X$ that assigns a (possibly structured) space D_X to each open subset $X \subset \mathcal{M}$

- For each X, Y with $Y \subset X$, a restriction map $x \to x|_Y$ from D_X to D_Y, such that $(x|_Y)|_Z = x|_Z$ (hence the notation $x|_Y$ is consistent without needing to mention what D_X x belongs to).

- For each $X \subset \mathcal{M}$, a map $d_X : \mathcal{F}_X \to D_X$, which commutes with restriction:

$$(d_X(\varphi))|_Y = d_Y(\varphi|_Y). \tag{15}$$

The idea is that d does not just describe the entire field: for any X, it describes the field *on* X. Clearly any local descriptor is a descriptor.

We define a local descriptor as:

- *separable* iff, given any collection of sets $\{X_i\}$ with $\cup_i X_i = X$, if $d_{X_i}(\varphi|_{X_i}) = d_{X_i}(\varphi'|_{X_i})$ then $d_X(\varphi|_X) = d_X(\varphi'|_X)$. In other words, a local descriptor is separable iff its description of a field on X depends only on its descriptions of the restrictions of the field to arbitrarily small elements of a partition of X.

- *Locally complete* iff for any $X \subset \mathcal{M}$, and any fields $\varphi, \varphi' \in \mathcal{F}_X$, if $d_X(\varphi) = d_X(\varphi')$ then φ and φ' are gauge-equivalent.

We can apply this framework to the usual gauge-invariant descriptors used in discussions of the A-B effect. The map $A \to F$ is a local descriptor, assigning to each gauge potential on a region its field strength on that same region; it is separable, but not locally complete, even though it is complete (as long as we continue to ignore matter fields) on simply connected manifolds. That is: knowing F everywhere on a simply connected space fixes the connection on that space up to gauge equivalence, but knowing F on a multiply-connected subset of that space does not fix the connection on that subset, even up to gauge equivalence. Conversely, the assignment to each $X \subset \mathcal{M}$ of the holonomies of curves in X is locally complete, but not separable.

As for gauge-fixings: the Coulomb gauge is not a local descriptor at all, despite its deceptively field-theoretic format: we cannot gauge-transform the restriction of a field to $X \subset \mathcal{M}$ into Coulomb gauge without knowing the gauge-invariant features of the field everywhere, not just in X, as we have seen. Nor are the Lorenz and R_ξ gauges, for similar reasons.

How general is this? It is straightforward to see that a large class of gauge-fixings fail to be local descriptors. Specifically, consider *local* gauge-fixings, in which the gauge is fixed by requiring

$$\mathcal{H}(x, A_\mu x, \partial_\nu A_\mu(x), \ldots \psi(x), \partial_\mu \psi(x), \ldots) = 0 \tag{16}$$

for all x and for some function $\mathcal{H}$, with the ... indicating possible dependence on finitely many higher derivatives. (All gauge-fixings I know are local gauge fixings; it is hard to see how any other gauge-fixing could provide a local descriptor, though I know of no rigorous results here.) Suppose that (ψ, A) is a field and that Λ is a gauge transformation which transforms it into a field that satisfies the gauge-fixing condition, and suppose also (which is generically the case) that the gauge group acts freely on (ψ, A), i. e. no nontrivial gauge transformation leaves (ψ, A) invariant, so that Λ is unique. We will then have some

equation for Λ, of form

$$\mathcal{H}(x, A_\mu(x) + \partial_\mu\Lambda(x), \partial_\nu(A_\mu(x) - \partial_\mu\Lambda(x)), \ldots, \psi'(x)\mathrm{e}^{ie\Lambda(x)}, \partial_\mu(\psi'(x)\mathrm{e}^{ie\Lambda(x)}), \ldots) = 0. \tag{17}$$

It is clear that in general this will be a *differential* equation, which suffices to uniquely determine Λ within X only given boundary conditions on X. For the gauge-fixing to be well-defined on $\mathcal{M}$ we will have to assume that the conditions (16) are supplemented by some boundary conditions on $\mathcal{M}$. But these do not in general determine boundary conditions on some subset $X \subset \mathcal{M}$: in general, even if we can solve the equations inwards from the boundary of $\mathcal{M}$ (which is by no means guaranteed) the boundary data on X will depend not only on the boundary conditions at $\mathcal{M}$ but also on the fields on the complement of X. Hence, two fields that are gauge-equivalent on X will in general not be gauge-fixed into fields that are equal on X.

The only way around this requirement is for $\mathcal{H}$ to depend at most on $\psi(x)$ and x, but not on $A(x)$ or any derivatives of ψ or A: only in this case will the gauge condition be an *algebraic* constraint on Λ. Since the modulus of ψ is gauge-invariant, this effectively makes the gauge condition into a condition on the phase of ψ: writing $\psi(x) = \rho(x)\mathrm{e}^{ie\phi(x)}$, the gauge condition becomes

$$\phi(x) = f(x, \rho) \tag{18}$$

for an arbitrary smooth function $f(x, \rho)$. Call this gauge-fixing condition *generalized unitary gauge* (fairly clearly it will almost always be simpler in practical applications just to use unitary gauge).

4   The limitations of unitary gauge

To sum up: most gauge-fixings provide a gauge-invariant description of a theory, but in most cases that description is highly nonlocal, describing the fields as a whole but providing no reliable insight into the spatial structure of those fields' gauge-invariant features. Unitary gauge (and its generalizations) are an exception, providing a description of the fields that is gauge-invariant, local, and separable.

Furthermore, unitary gauge has a natural interpretation in terms of directly-physical, obviously gauge-invariant features of a theory. The (always-real, always-positive) field ψ in unitary gauge represents the manifestly gauge-invariant quantity $|\psi|$. (In a generalized unitary gauge, this is represented by $\psi\mathrm{e}^{-ief(x,|\psi|)}$). The connection A satisfies the equation

$$\mathcal{D}_\mu\psi = \partial_\mu\psi + ieA_\mu\psi \tag{19}$$

and in unitary gauge ψ is real, so that $A_\mu\psi$, up to charge, is the imaginary part of the covariant derivative. If we write $\psi = \rho\mathrm{e}^{ie\phi}$, this can be rewritten as

$$A_\mu = \partial_\mu\phi \tag{20}$$

so that A just represents the (again, manifestly gauge-invariant) covariant derivative of the phase. (See (Struyve 2011) for further consideration of this close relationship between unitary gauge and a manifestly gauge-invariant description.)

This might suggest a preferred status for (generalized) unitary gauge: it, and it alone, provides a local, separable, representation of the gauge-invariant features of a field. There are two reasons to be cautious of this conclusion, which I will refer to as the Degeneracy Reason and the Incompleteness Reason.

The Degeneracy Reason is that unitary gauge is not actually a gauge-fixing whenever $\psi = 0$ on an open subset of $\mathcal{M}$. For then we can assign the phase of the scalar field arbitrarily, and more importantly, then we cannot recover the connection (and hence the gauge-invariant field strength) from the (vanishing) covariant derivative of ψ.

There is something important to this objection (to which I return below) but in itself it should not be taken seriously. If ψ *literally vanishes* on an open set X, then unitary gauge is ill-defined on X. But if $0 < |\psi| < 10^{-10^{100}}$kg on X, unitary gauge is perfectly well defined.[5] And to suppose that it is physically significant whether $|\psi| = 0$ or merely $|\psi| < 10^{-10^{100}}$kg is to take scalar electrodynamics far more literally than it has any right to be taken. It is an approximately-accurate description of the structure of certain systems on certain scales, not a candidate for the Theory of Everything.

The point is only sharpened in quantum field theory, where classical fields are not direct descriptions of a system but merely variables in a path integral. That path integral requires regularization to be well-defined, but under any reasonable regulararization, fields that vanish on an open set will have measure zero, and so will not contribute to the path integral.

That said, the Degeneracy Reason does suggest that unitary gauge will be misleading or unhelpful — even if strictly correct — when applied to regions where the matter field is negligible. One way to characterize the abstract structure of a gauge theory is that a choice of gauge corresponds to a reference frame (for scalar electrodynamics, a choice of the real axis) in the internal field space at each point in space, and that the connection tells us how to relate those reference frames along any given path. Unitary gauge amounts to letting the matter field itself determine the reference frame; this has clear physical significance when the matter field is dynamically relevant, but seems rather arbitrary when it is negligible.

Similarly, if we are using quantum-field-theoretic methods to study electrodynamics in a region where there are no matter particles, it is natural to integrate out the matter field entirely, obtaining an effective field theory for the electromagnetic field alone. The resulting theory will have only A as its degree of freedom, and so we cannot define unitary gauge. (It remains well-defined in the full quantum field theory, but the degrees of freedom it uses are dynamically irrelevant in this regime.)

The Incompleteness Reason is more compelling, and applies whether or not

---

[5]$|\psi|$ has the dimensions of mass in units where $\hbar$=1. If you prefer SI units, use them; manifestly the point I make here is independent of a choice of scale.

the matter field is negligible. Consider a constant phase shift, $\psi(x) \to \psi \mathrm{e}^{ie\theta}$, where θ is spatially independent. Manifestly, the unitary-gauge description of a field is invariant under that transformation: it leaves both A and $|\psi|$ invariant. But the physical features of the system are *not* invariant, for this transformation is not a small gauge transformation: it is a physical shift of the system, representing a change in its phase as measured by an observer outside the system. This means that, in the terminology of section 1, unitary gauge is not a gauge-fixing at all, since it does not assign an element to every gauge equivalence class: if the phase of ψ tends to anything other than zero at infinity, there is no small gauge transformation which puts ψ in unitary gauge. Put another way: putting a field into unitary gauge discards genuine information — the average phase of the scalar field — along with descriptive fluff. If the Degeneracy Reason gives cause to worry that unitary gauge is not a gauge-invariant descriptor, the Incompleteness reason gives cause to worry that it is not a complete descriptor.

(Similarly, if a gauge theory is defined on a topologically non-trivial manifold, the gauge transformation required to put a field in unitary gauge may be topologically large; this occurs in the solenoid spacetime of section 1. Again, this means that unitary gauge conflates two fields that are distinguished one from another by their relation with external systems, though the nature of that relation is subtler here — see (Wallace 2022b).)

It is helpful here to compare gauge theory to Newtonian mechanics. In the latter theory, it is fairly straightforward (Barbour and Bertotti 1982; Barbour 2012) to start with the dynamics of an isolated system and then quotient out its translational and rotational symmetries; the result is a self-contained dynamical theory of how the intrinsic features of the theory change with time. But doing so throws away genuine information, since (at least in most applications) Newtonian mechanics describes not a fictional system in an otherwise-empty universe but a *de facto* isolated subsystem of a larger cosmos, and symmetry-related states of that system are differently located and oriented with respect to that larger cosmos (Wallace 2022a). In the same sense, the unitary-gauge description of a field can helpfully be understood as representing the intrinsic features of an isolated subsystem. A single additional, nonlocal, piece of information — its average phase — is required to specify the system completely, at least if it too is understood as an isolated subsystem in a larger cosmos.

But the Incompleteness Reason has an important limitation: it assumes that global phase shifts actually are symmetries of the theory. Whether this is true depends sensitively on the boundary conditions of the theory. If that boundary condition imposes $\lim_{|\vec{x}| \to \infty} \psi = 0$, it is preserved by global phase shifts, and so ψ and $\psi \mathrm{e}^{ie\theta}$, for constant θ, are distinct possible states. But if it imposes a nonzero infinite limit ψ_0 for ψ, that limit will not be preserved under global phase shifts and so they will not represent a physical symmetry. In the latter, but not the former, case, (generalized) unitary gauge is indeed a gauge-fixing, and moving to (generalized) unitary gauge provides a lossless description of a field's gauge-invariant properties.

The distinction between these two scenarios can be usefully illustrated by

considering the spectrum of elementary excitations of the theory, which in classical field theory is the space of solutions of the linearized field equations, and in quantum field theory is the space of particles. In either case we obtain it by looking at the quadratic part of the expansion of the action of the theory around its minimum. In classical field theory the action is given once and for all and determines the classical equations of motion; in quantum field theory the action is the (temperature-dependent) effective action and its successive derivatives determine the vacuum expectation value and the one-particle-irreducible (1PI) N-point functions, with the quadratic part of the expansion determining the (inverse of the connected part of the) 2-point function and hence the mass spectrum.

In the case of an unbroken symmetry (boundary conditions where the matter field vanishes) the elementary excitations consist of two excitations of the matter field, related by a 90-degree phase shift (the signature of this in QFT is antimatter; cf (Baker and Halvorson 2009; Wallace 2009)) and describable by a complex scalar field, and two excitations of the gauge connection, forming a massless vector field (the exact description of which varies from gauge to gauge). Here the existence of multiple phase-related modes of the matter field is entirely hidden in unitary gauge in the quadratic expansion (of course it can be seen at higher order points in the expansion) while being manifest in Lorenz, Coulomb, or R_ξ gauge.

The awkwardness of accessing the spectrum from unitary gauge can be traced back to the Degeneracy Reason: we are trying to expand the field around $\phi = 0$, but unitary gauge fails on that configuration, and so the expansion is badly behaved.

If the symmetry is broken (in other words, if the boundary conditions are such that the matter field is nonvanishing) we can safely use unitary gauge to study the excitation spectrum; here, as any elementary QFT textbook will tell you, we encounter the Higgs mechanism, where the excitations comprise a single excitation of the matter field describable by a real scalar field, and three excitations of the gauge connection which jointly form a massive vector field (and which are described in an entirely local, gauge-invariant manner if we gauge-fix to unitary gauge).

(What becomes of the analogy to Newtonian mechanics in this case? The short answer is that if the symmetry is spontaneously broken, we cannot simply combine isolated systems with different boundary conditions to make a larger system. For the longer answer, see (Wallace 2022b).)

I pause to note two subtleties. Firstly, in classical field theory whether a symmetry is broken or not is a once-and-for-all feature of a theory: it is fixed by the action, and the action encodes the fundamental dynamics. In quantum field theory, by contrast, the action is an effective function determined by the temperature, and the same theory can display different symmetry-breaking phenomena at different temperatures. It thus makes literal sense in QFT (but only metaphorical sense in a classical theory) to talk about symmetry-broken and symmetry-unbroken *sectors* of the same theory.

Secondly, this attribution of excitations to matter rather than gauge con-

nection, or vice versa, should be taken with a grain of salt, since it is not gauge invariant. (We can study the symmetry-broken sector of scalar electrodynamics in R_ξ gauge perfectly well; if we are feeling masochistic we can even study it in Coulomb or Lorenz gauge.) It is of course gauge-invariant to say that *in unitary gauge* the elementary excitations of the symmetry-broken sectors comprise one matter and two gauge-connection excitations.

5   Locality and unitary gauge in non-Abelian gauge theories

(This section lies outside the main flow of the paper, and may be skipped on first reading.)

The general arguments of section 3 apply to non-Abelian gauge theories too: in order to be a local descriptor, a gauge-fixing must algebraically determine the gauge transformation required to take any field into its gauge-fixed form, and this means it must fix the gauge-variant features of the matter field. For the moment, consider this problem for a scalar matter field, taking values in some space $\mathcal{V}$ on which the gauge group G acts. We can partition $\mathcal{V}$ into *orbits*, equivalence classes under the action of $\mathcal{V}$; the generalization of unitary gauge requires us to pick, in each orbit, some representative element.[6] Let $f : \mathcal{V} \to \mathcal{V}$ map each point in $\mathcal{V}$ to the representative element in its orbit; then we gauge-fix a field (ψ, A) by applying a gauge transformation u such that $u(x)\psi(x) = f(\psi(x))$. (For scalar electrodynamics, f is given by $f(\psi) = |\psi|$.)

Does this actually determine a gauge-fixing? Only if it fixes u uniquely, which requires that no gauge transformation u' distinct from u exists such that $u(x)\psi(x) = u'(x)\psi(x)$, which is equivalent to the requirement that no non-trivial gauge transformation v exists such that $v(x)\psi(x) = \psi(x)$ (for any x). In other words, the group G should act *freely* on every orbit of G in $\mathcal{V}$ in which $\psi(x)$ takes values. Requiring gauge transformations to be smooth weakens this slightly: there must be no orbit of G in $\mathcal{V}$ such that (i) G does not act freely on $\mathcal{V}$; (ii) there is an open set in spacetime such that ϕ, restricted to that set, takes values in the orbit.

For scalar electrodynamics, there is only one such non-free orbit: the point $\psi = 0$, already discussed. But in many scalar non-Abelian theories, there will be many. Take, for instance, the standard action of $SU(N)$ or $U(N)$ on $\mathbb{C}^N$. The orbits are the (complex) spheres of fixed $|\psi|$, and $SU(N)$ does not act freely on any such orbit: for any nonzero vector in $\mathbb{C}^N$, there will be a $U(1)$ subgroup of $SU(N)$ that leaves it invariant. (This is essentially the mathematics behind the breakdown of electroweak $SU(2) \times U(1) \equiv U(2)$ symmetry to $U(1)$ symmetry in the Standard Model.)

So unitary gauge is not well-defined for non-Abelian gauge theories with scalar matter, at least in these important cases. But in general it will be defin-

---

[6]We could choose to let the choice of representative element depend on spatial location, but nothing is gained by such generalization.

able for more complicated (and realistic) matter fields. If ψ is a spinor matter field, for instance, $\psi(x)$ can be thought of as a linear map from $\mathbb{C}^2$ to $\mathcal{V}$, with the gauge group acting on the space of such maps; generic points in the space of such maps transform freely under gauge transformations, so that specifying a representative element in each orbit will fix the gauge of any field except for some non-generic ones. Unitary gauge then works as well for non-Abelian as for Abelian theories: that is, it works, up to the Degeneracy and Incompleteness reasons, which apply to the non-Abelian case *mutatis mutandis.*

It must be admitted, though, that there is something awkward and cumbersome about these generalizations of unitary gauge (and indeed I am not aware of any such generalization having applications in the physics literature). We can understand why this is by considering again the interpretation of unitary gauge I offered in the previous section: a choice of gauge amounts to a choice of reference frame at each point; unitary gauge uses the matter field itself to fix a choice of reference frame. For scalar electrodynamics there is an obvious best way to do so: any nonzero field at a point has a phase and the connection along a path between any two points can be interpreted simply and invariantly as giving the change in phase between the fields at those two points. When the matter field is more complicated, there may be no obviously most natural rule to determine a reference frame or to define the transformation between two such frames, even when the field has sufficient structure that such rules can be stated. (Similar issues would arise even in electrodynamics given a more complicated, e.g. spinor, matter field.)

6   Conclusion

Gauge-invariant descriptions of a theory with local gauge symmetry are easier to come by than you may think: any gauge-fixing for a theory provides a gauge-invariant description of that theory's fields, just as any description of a spatial system in an antecedently-specified coordinate system provides a coordinate-invariant description of that spatial system.

However, most such descriptions are deceptive. On the surface they appear to be a *local* description of the system (specified as they are in terms of the values of local fields), but in reality they are generally nonlocal, in the sense that two fields gauge-equivalent in a region can often have distinct gauge-fixings in that reason.

The exception is unitary gauge (along with its generalizations). Unitary gauge is local (thus improving on most other gauges), gauge-invariant (thus improving on a description in terms of field strengths) and separable (thus improving on a description in terms of holonomies), and explicitly includes matter degrees of freedom (which field-strength and holonomy descriptions do not). However, in the unbroken sector of a gauge theory, unitary gauge provides an incomplete description of the system (although a complete description of the intrinsic features of the system), since it leaves out overall phase and potentially also topological information; in cases more complicated than scalar electrody-

namics it may also be quite awkward, and there may be no natural reason to choose one version of unitary gauge over another.

What does this mean for the search for gauge-invariant descriptions of physical processes? In a sense they are easy to come by: pick a gauge, describe the physics in that gauge, and call it a day. But the nonlocal aspects of most gauges may make the description thus obtained (gauge-invariant, but) somewhat opaque. In many cases we can avoid this problem by using unitary gauge, but this must be done with some caution if asymptotic symmetry is unbroken, since that description is incomplete (although it ought still to be adequate if all that is desired is a description of the system's intrinsic behavior), obscures the particle spectrum, and is awkwardly suited to describing situations where the matter field is negligible.

The Higgs mechanism provides a brief illustration. Philosophers of physics have long[7] sought a gauge-invariant description of it: here is one (closely related to that offered by (Struyve 2011)). Firstly, we must ask whether asymptotic global symmetry is broken in the sector of the theory being studied: this can be characterized simply and gauge-invariantly, and its 'tell' is the nonvanishing expectation value of the magnitude of the Higgs field in the ground state (or at thermal equilibrium in a finite-temperature sector). If it is unbroken, the Higgs mechanism does not apply, and the symmetry is manifest in the particle spectrum. If it is broken, unitary gauge offers a complete, gauge-invariant, local, separable description of the physics, in which it is easy to see that the vector boson field is massive and that there is no Goldstone mode.

I close with remarks aimed at a somewhat different audience. I have argued that gauge fixings offer gauge-invariant descriptions of the physics, so that different gauge-fixings offer different gauge-invariant descriptions. But which description is best? The question has an obvious pragmatic interpretation: 'best' means 'most helpful for the purpose at hand', in which case we should adopt whichever choice we find most helpful (even a nonlocal choice), with little care as to the rightness of that choice beyond its use.

But there is a tradition in the metaphysics of physics (illustrated very clearly by (Sider 2020)), where the question is rather: which description is *most fundamental*, which description *carves nature at its joints*? For reasons I explicate in (Wallace 2022d), I am sceptical that there is a meaningful question here once we set pragmatism aside. For those who disagree, I recommend unitary gauge as the clearly-preferable choice, notwithstanding its limitations when applied to subsystems of the larger cosmos — an issue which in any case I suspect will not bother those whose metametaphysics is less deflationary than mine. But I also note that the range of possible versions of unitary gauge for systems more complicated than scalar electrodynamics may ultimately cause metaphysical difficulties for any such strategy.

---

[7]See, e. g. , (Earman 2003; Smeenk 2006; Lyre 2008).

Acknowledgements

References

Baker, D. and H. Halvorson (2009). Antimatter. *British Journal for the Philosophy of Science 60*, 1–29.

Barbour, J. (2012). Shape dynamics: an introduction. In F. Finster, O. Müller, M. Nardmann, J. Tolksdorf, and E. Zeidler (Eds.), *Quantum Field Theory and Gravity: Conceptual and Mathematical Advances in the Search for a Unified Framework*, pp. 257–297. Springer. Available online at http://arxiv.org/abs/1105.0183.

Barbour, J. and B. Bertotti (1982). Mach's principle and the structure of dynamical theories. *Proceedings of the Royal Society of London 382*, 295.

Belot, G. (2018). Fifty million Elvis fans can't be wrong. *Nous 52*, 946–981.

Cheng, T.-P. and L.-F. Li (1984). *Gauge Theory of Elementary Particle Physics*. Oxford, UK: Oxford University Press.

Earman, J. (2003). Curie's principle and spontaneous symmetry breaking. *International Studies in the Philosophy of Science 18*, 335–346.

Healey, R. (1997). Nonlocality and the Aharonov-Bohm effect. *Philosophy of Science 4*, 18–41.

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

Lyre, H. (2008). Does the Higgs mechanism exist? *International Studies in the Philosophy of Science 22*, 119–133.

Peskin, M. E. and D. V. Schroeder (1995). *An introduction to Quantum Field Theory*. Reading, Massachusetts: Addison-Wesley.

Sider, T. (2020). *The Tools of Metaphysics and the Metaphysics of Science*. Oxford: Oxford University Press.

Smeenk, C. (2006). The elusive Higgs mechanism. *Philosophy of Science 73*, 487–499.

Struyve (2011). Gauge-invariant accounts of the Higgs mechanism. *Studies in the History and Philosophy of Modern Physics 42*, 226–236.

Wallace, D. (2009). QFT, antimatter, and symmetry. *Studies in the History and Philosophy of Modern Physics 40*, 209–222.

Wallace, D. (2014). Deflating the Aharonov-Bohm effect. Forthcoming; available online at https://arxiv.org/abs/1407.5073.

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

Wallace, D. (2022a). Isolated systems and their symmetries, part I: General framework and particle-mechanics examples. *Studies in the History and Philosophy of Science 92*, 239–248.

Wallace, D. (2022b). Isolated systems and their symmetries, part II: Local and global symmetries of field theories. *Studies in the History and Philosophy of Science 92*, 249–259.

Wallace, D. (2022c). Observability, redundancy and modality for dynamical symmetry transformations. In J. Read, B. Roberts, and N. Teh (Eds.), *The Philosophy and Physics of Noether's Theorems: A Centenary Volume*, pp. 322–353. Cambridge: Cambridge University Press.

Wallace, D. (2022d). Stating structural realism: mathematics-first approaches to physics and metaphysics. In J. Hawthorne (Ed.), *Philosophical Perspectives Volume 36: Metaphysics*, pp. 345–378. Wiley-Blackwell.

Weatherall, J. (2016). Understanding gauge. *Philosophy of Science 83*, 1039–1049.

Weinberg, S. (1995). *The Quantum Theory of Fields, Volume II: Modern Applications*. Cambridge: Cambridge University Press.