The Aharonov-Bohm Effect Explained: Reality of Gauge Potentials and Its Implications

Shan Gao
Research Center for Philosophy of Science and Technology,
Shanxi University, Taiyuan 030006, P. R. China
E-mail: gaoshan2017@sxu.edu.cn.

March 19, 2025

Abstract

The Aharonov-Bohm (AB) effect, wherein an electron acquires a phase shift in a field-free region due to electromagnetic potentials, poses a profound challenge to the ontology of quantum mechanics and gauge theories. This paper demonstrates that gauge-invariant explanations, which attribute the phase to measurable quantities, rely on nonlocal and discontinuous mechanism and fail to account for its continuous accumulation along the electron's path—a process evident in the generalized AB effect, where a time-varying magnetic flux induces a phase that builds gradually over time, as predicted by quantum mechanics. Through a new analysis spanning quantum mechanics and quantum electrodynamics, I argue that the electromagnetic potentials A_μ, fixed in the Lorenz gauge, emerge as the fundamental physical reality, offering a local, relativistically consistent account of the phase's generation. This exclusion of gauge-invariant paradigms reverberates across gauge theories: it redefines the Higgs mechanism, favoring dynamic potentials over static invariants, and extends to general relativity, where gravitational potentials $g_{\mu\nu}$ may anchor spacetime's substantival reality via a gravitational AB effect. By unraveling the AB phase's continuous generation—locally mediated by potentials—this study not only addresses a long-standing conundrum but also bridges electromagnetic, gravitational, and Yang-Mills frameworks, offering a unified potential-centric perspective with novel implications for physics and philosophy.

The eternal mystery of the world is its comprehensibility.
—Albert Einstein, *Physics and Reality*, 1936

Contents

1 Introduction **4**

2 The AB Effect and Its Generalized Form **5**
2.1 The Magnetic AB Effect 5
2.2 Generalized Magnetic AB Effect with Time-Varying Flux . . 6
2.3 Electric AB Effect with Time-Varying Scalar Potential 8

3 Explanations in Terms of Gauge-Invariant Quantities **9**
3.1 Why Gauge-Invariant Explanations Appear Preferable 9
3.2 Gauge-Invariant Quantities in Quantum Mechanics 10
3.3 Dynamics Before and After Overlapping 12
3.4 Critique of Gauge-Invariant Explanations 12
3.5 Madelung Hydrodynamics as a Paradigm 14
3.5.1 Derivation of the Madelung Equations 14
3.5.2 Inequivalence to the Schrödinger Equation 15
3.5.3 Nonlocality of the Quantization Condition 15
3.5.4 Failure to Explain the AB Effect 16
3.6 Healey's Loop Ontology and Its Objections 16
3.6.1 Dynamical Inadequacy of the Loop Ontology 17
3.6.2 Structural Redundancy in the Loop Framework 17
3.6.3 Nonlocal Interdependence and Causal Concerns 17

4 A No-Go Result for Gauge-Invariant Explanations **18**
4.1 An Incompatibility Proof 18
4.2 Objections and Rebuttals 21
4.3 Experimental Proposals 21
4.4 Conclusion 23

5 Is Quantum Electrodynamics Relevant? **24**
5.1 Correspondence Between QM and QED 24
5.2 Gauge Invariance in a QED Framework 25
5.3 Vaidman's Alternative Proposal 25
5.4 Implications for Gauge-Invariant Explanations 25
5.5 Conclusion 26

6 Reality of potentials in the Lorenz gauge **26**
6.1 The Physical Significance of Electromagnetic Potentials . . . 26
6.2 The Necessity of One True Gauge 28
6.3 The Lorenz Gauge as the True Gauge 29
6.3.1 Locality and Causal Propagation 30
6.3.2 Lorentz Covariance 30
6.3.3 Determinacy and Uniqueness 30
6.3.4 Ontological Sufficiency 31

6.4 Nonzero Photon Mass and the Proca Equation 31
6.5 Conclusion 32

7 The Ontological Status of Potentials in the Lorenz Gauge 33
7.1 Potentials as the Electromagnetic State 33
7.2 The Photon Wave Function 33
7.2.1 Definition of the Photon Wave Function 34
7.2.2 Proof of Completeness and Uniqueness 35
7.2.3 Explaining the Double-Slit Experiment as an Example 36
7.2.4 Comparison with Alternatives 37
7.2.5 Conclusion 39
7.3 Quantum Origin of Classical Potentials 40

8 Implications for Gauge Theories 41
8.1 The Origin of the Berry Phase 41
8.2 Non-Abelian Gauge Theories 42
8.2.1 Non-Abelian Gauge Structure and Potentials 43
8.2.2 Non-Abelian Analogs of the AB Effect 43
8.2.3 Nonlocality and Discontinuity Concerns 44
8.2.4 Incompleteness and Potential-Centric Alternatives . . 44
8.2.5 Broader Gauge Theory Implications 44
8.3 The Higgs Mechanism 45
8.3.1 Overview of the Higgs Mechanism and its Gauge-Invariant Reformulation 46
8.3.2 Problems of the Gauge-Invariant Reformulation 47
8.3.3 Analogy Between the AB Effect and the Higgs Mechanism 49
8.3.4 Challenges of the Unitary Gauge in Defect Configurations 51
8.3.5 The True Gauge for the Higgs Mechanism 53
8.3.6 The Physical Meaning of Gauge Symmetry Breaking . 55
8.4 General Relativity and Gravitational Potentials 58
8.4.1 The Hole Argument and the Quest for Spacetime's Reality 58
8.4.2 Gravitational AB Effects and the Substantivalist Turn 59
8.4.3 Ontological Implications and Philosophical Horizons . 60
8.4.4 Massive Gravity: A Proca-like Path to Gravitational Determinacy 60
8.5 Beauty in the Unseen 62

9 Conclusions and Future Directions 62

1 Introduction

The Aharonov-Bohm (AB) effect is one of the most intriguing phenomena in quantum mechanics, challenging our classical intuitions about the nature of physical interactions. Conceived by Werner Ehrenberg and Raymond E. Siday in 1949 and fully articulated by Yakir Aharonov and David Bohm in 1959, it demonstrates that an electron, traversing a region free of electric and magnetic fields, yet permeated by electromagnetic potentials, acquires a phase shift observable through interference upon recombination of bifurcated beams. At the heart of this phenomenon lies a vexing question: how does a gauge-invariant phase shift emerge during the electron's journey through a field-free region, and what does this imply about the nature of reality in quantum mechanics? This enduring puzzle, defying full explanation within traditional frameworks (Earman, 2024; Wallace, 2024), anchors this study's exploration of quantum ontology and gauge theories.

Gauge-invariant accounts, rooted in classical intuition, assert that only measurable quantities like field strengths hold physical reality, dismissing potentials as mathematical artifacts. Yet, these perspectives stumble on the phase's genesis: how does a gauge-invariant shift arise in a field-free transit? This paper argues that the AB effect demands a reconceptualization, positing electromagnetic potentials as ontologically primary, a view bolstered by the generalized AB effect's dynamic phase accumulation. Through a rigorous critique spanning quantum mechanics (QM), quantum electrodynamics (QED), and gauge theories, it challenges the adequacy of gauge-invariant paradigms, advocating a potential-centric ontology fixed in the Lorenz gauge.

The exploration unfolds across nine sections, weaving empirical evidence, theoretical rigor, and philosophical reflection to resolve the AB effect's enigma and extend its lessons across physics. Section 2 lays the groundwork by detailing the standard and generalized AB effects, revealing the phase's continuous accrual as a cornerstone of quantum reality. Section 3 dissects gauge-invariant explanations—spanning field-based, Madelung hydrodynamic, and topological loop ontologies—exposing their nonlocal and discontinuous flaws. Section 4 delivers a decisive blow, leveraging the generalized AB effect to exclude these accounts, supported by experimental proposals to test the phase's dynamic genesis. Section 5 probes quantum electrodynamics' reinforcement of this critique, affirming potentials' primacy over quantized invariants. Section 6 establishes electromagnetic potentials A_μ in the Lorenz gauge as the true reality, detailing its significance (6.1), the need for a privileged gauge (6.2–6.3), and a speculative extension to massive photons via the Proca equation (6.4). Section 7 explores the ontological status of potentials, casting them as the electromagnetic state (7.1), defining the photon wave function (7.2), and tracing classical potentials' quantum origins (7.3). Section 8 casts this potential-centric ontology across gauge theories: reinterpreting the Berry phase's origins (8.1), challenging non-Abelian Yang-Mills frameworks (8.2), critiquing the Higgs mechanism's invariant reformulations (8.3), and proposing a gravitational AB effect to affirm spacetime's substantival essence in general relativity (8.4). Section 9 synthesizes these results, outlining future experimental, theoretical, and philosophical paths to cement the AB effect as a fulcrum for reshaping our understanding of physical reality.

2 The AB Effect and Its Generalized Form

The AB effect, once a theoretical curiosity, has been experimentally validated (Tonomura et al, 1986) and mathematically refined over decades (Ballesteros and Weder, 2009). This section outlines the standard magnetic AB effect (Section 2.1), its generalized form with time-varying flux (Section 2.2), and the electric AB effect with time-varying scalar potential (Section 2.3), laying the empirical and theoretical groundwork for subsequent critiques of gauge-invariant explanations and the advocacy of a potential-centric ontology. Central to these phenomena is the continuous accumulation of phase over time, a feature that challenges traditional field-centric views and unifies the magnetic and electric AB effects under a dynamic potential framework.

2.1 The Magnetic AB Effect

Consider a canonical experiment: a coherent electron beam is emitted, split into two paths encircling an infinitely long solenoid, and recombined to form an interference pattern. A steady current within the solenoid generates a magnetic field $\mathbf{B} = \nabla \times \mathbf{A}$, confined to its interior, ensuring the exterior region traversed by the electron beams is devoid of magnetic fields ($\mathbf{B} = 0$). Yet, the vector potential $\mathbf{A}$ persists outside because the magnetic flux $\Phi = \oint \mathbf{B} \cdot d\mathbf{s} = \oint \mathbf{A} \cdot d\mathbf{r}$ through any loop enclosing the solenoid is nonzero. This subsection derives the phase shift for this static scenario, establishing the foundational result characteristic of the standard magnetic AB effect.

The Schrödinger equation for an electron in this setup (with $\hbar = c = 1$) is:

$$i\frac{\partial \psi}{\partial t} = -\frac{1}{2m}(\nabla - ie\mathbf{A})^2\psi + eA_0\psi, \tag{1}$$

where e and m are the electron's charge and mass, $\mathbf{A}$ is the magnetic vector potential, and A_0 is the electric scalar potential, set to zero ($A_0 = 0$) outside the solenoid. This equation exhibits gauge invariance under the transformations:

$$A_\mu \to A'_\mu = A_\mu - \partial_\mu \Lambda, \quad \psi \to \psi' = e^{-ie\Lambda}\psi, \tag{2}$$

where $A_\mu = (-A_0, \mathbf{A})$ and $\Lambda(x)$ is a smooth gauge function. In a simply connected region, choosing:

$$\Lambda(x) = \int_{x_0^\mu}^{x^\mu} A_\mu(x')dx'^\mu, \tag{3}$$

transforms A_μ to zero, yielding $\psi = e^{ie\Lambda}\psi_0$, where ψ_0 satisfies the free Schrödinger equation. Consequently, the phase accumulated along a path L due to the potential is:

$$\Delta\phi = e\int_L A_\mu dx^\mu. \tag{4}$$

In the standard magnetic AB effect with a constant flux Φ, the phase difference between the two paths L_1 and L_2, forming a closed loop $C = L_1 - L_2$ around the solenoid, is:

$$\phi_{AB} = e\int_{L_1} \mathbf{A}\cdot d\mathbf{r} - e\int_{L_2} \mathbf{A}\cdot d\mathbf{r} = e\oint_C \mathbf{A}\cdot d\mathbf{r} = e\Phi \tag{5}$$

by Stokes' theorem. This phase shift, observable as a shift in the interference pattern, arises despite the electron experiencing no local magnetic forces, highlighting the potential's physical significance in the static case. The result $\phi_{AB} = e\Phi$ is the hallmark of the standard magnetic AB effect, where the flux's constancy ensures a straightforward, time-invariant contribution to the quantum phase, setting the stage for the dynamic generalization explored in the subsequent subsection.

2.2 Generalized Magnetic AB Effect with Time-Varying Flux

When the magnetic flux $\Phi(t)$ varies with time, the AB effect evolves into a dynamic phenomenon, requiring a detailed analysis of the electron's motion under a time-dependent magnetic vector potential. Unlike the static case, the phase shift reflects the temporal profile of $\Phi(t)$, necessitating consideration of the electron's trajectory and the induced electric field. Here, I derive the phase shift, building on the gauge-invariant framework of Section 2.1, and show that it is proportional to the time-averaged enclosed magnetic flux over the transit duration, as detailed in Gao (2025a).

Consider the AB setup: a coherent electron beam is emitted at $t = 0$, split into two paths encircling an infinitely long solenoid, and recombined at $t = T$. The solenoid carries a time-varying magnetic flux $\Phi(t)$, with the magnetic field $\mathbf{B} = \nabla\times\mathbf{A}$ confined within, maintaining $\mathbf{B} = 0$ outside. In a gauge where $A_0(\mathbf{r}, t) = 0$, the vector potential outside is:

$$\mathbf{A}(\mathbf{r}, t) = \frac{\Phi(t)}{2\pi r}\hat{\boldsymbol{\theta}}, \tag{6}$$

where r is the radial distance and $\hat{\boldsymbol{\theta}}$ is the azimuthal unit vector, ensuring $\oint_C \mathbf{A} \cdot d\mathbf{r} = \Phi(t)$ for a closed loop C. The total phase shift is:

$$\phi_{AB} = e \int_{L_1} \mathbf{A}(\mathbf{r}, t) \cdot d\mathbf{r} - e \int_{L_2} \mathbf{A}(\mathbf{r}, t) \cdot d\mathbf{r} = e \oint_C \mathbf{A}(\mathbf{r}, t) \cdot d\mathbf{r}. \tag{7}$$

For circular paths of radius R, the displacement along each path is $d\mathbf{r} = \omega_i(t) R dt \hat{\boldsymbol{\theta}}$, where $\omega_i(t)$ (for $i = 1, 2$) is the angular velocity of each beam. Thus the phase shift becomes:

$$\phi_{AB} = e \int_0^T \frac{\Phi(t)}{2\pi} [\omega_1(t) + \omega_2(t)] dt, \tag{8}$$

where T is the transit time, and $\omega_1(t)$ and $\omega_2(t)$ are the angular velocities for the counterclockwise and clockwise paths, respectively, with $\int_0^T [\omega_1(t) + \omega_2(t)] dt = 2\pi$ for a complete encirclement.

The time-varying $\Phi(t)$ induces an electric field $\mathbf{E} = -\frac{\partial \mathbf{A}}{\partial t}$, calculated as:

$$\mathbf{E} = -\frac{1}{2\pi R} \frac{d\Phi(t)}{dt} \hat{\boldsymbol{\theta}}, \tag{9}$$

which exerts a tangential force $F_\theta = eE_\theta = -\frac{e}{2\pi R} \frac{d\Phi}{dt}$, accelerating one beam and decelerating the other. The equation of motion for each beam's angular velocity is:

$$mR \frac{d\omega_i}{dt} = F_\theta, \tag{10}$$

yielding:

$$\frac{d\omega_1}{dt} = -\frac{e}{2\pi m R^2} \frac{d\Phi}{dt}, \quad \frac{d\omega_2}{dt} = \frac{e}{2\pi m R^2} \frac{d\Phi}{dt}. \tag{11}$$

Integrating from $t = 0$ to t, with initial velocities $\omega_1(0)$ and $\omega_2(0)$:

$$\omega_1(t) = \omega_1(0) - \frac{e}{2\pi m R^2} [\Phi(t) - \Phi(0)], \tag{12}$$

$$\omega_2(t) = \omega_2(0) + \frac{e}{2\pi m R^2} [\Phi(t) - \Phi(0)]. \tag{13}$$

The combined angular velocity simplifies to:

$$\omega_1(t) + \omega_2(t) = \omega_1(0) + \omega_2(0), \tag{14}$$

as the induced field's effects cancel, and $\int_0^T [\omega_1(0) + \omega_2(0)] dt = 2\pi$ leads to $\omega_1(0) + \omega_2(0) = \frac{2\pi}{T}$. Substituting into Equation (8):

$$\phi_{AB} = \frac{e}{2\pi} \int_0^T \Phi(t) [\omega_1(0) + \omega_2(0)] dt = \frac{e}{2\pi} \cdot \frac{2\pi}{T} \int_0^T \Phi(t) dt = \frac{1}{T} \int_0^T e\Phi(t) dt. \tag{15}$$

This confirms the phase shift as the time average of $e\Phi(t)$, reducing to $\phi_{AB} = e\Phi$ when $\Phi(t) = \Phi$.

The generalized AB effect reveals a hybrid nature: the region outside the solenoid is not strictly field-free due to the induced electric field $\mathbf{E} = -\frac{\partial \mathbf{A}}{\partial t} = -\frac{1}{2\pi R}\frac{d\Phi(t)}{dt}\hat{\boldsymbol{\theta}}$, which is finite during flux variations. However, for a very short transition duration (e.g., $\Delta t \ll T$), $\mathbf{E}$'s influence on the phase, which is proportional to the duration, is negligible. Furthermore, the field's effects cancel in the phase difference: opposite accelerations on the two beams ensure $\phi_{AB} = \frac{1}{T}\int_0^T e\Phi(t)\,dt$ reflects only the time-averaged flux, not trajectory perturbations. This ensures that the generalized AB effect remains a clean probe of the vector potential $\mathbf{A}$'s influence, even in the presence of transient induced fields.

2.3 Electric AB Effect with Time-Varying Scalar Potential

Complementing its magnetic counterpart, the electric AB effect showcases the quantum influence of the scalar potential A_0 in regions where $\mathbf{E} = -\nabla A_0 - \partial \mathbf{A}/\partial t = 0$, encompassing both static and time-varying scenarios. The time-varying case mirrors the dynamic framework of the generalized magnetic AB effect (Section 2.2), emphasizing the continuous accumulation of phase over time.

Consider a setup where a coherent electron beam splits into two paths at $t = 0$, traverses field-free regions (e.g., within Faraday cages), and recombines at $t = T$. Along path L_2, a time-varying scalar potential $A_0(t)$ is applied, while path L_1 maintains $A_0 = 0$, with dynamic shielding ensuring $\mathbf{E} = 0$ and $\mathbf{B} = 0$ along both paths via $\mathbf{A}$ adjustments, as in Section 2.1. The phase shift along a path L is:

$$\Delta\phi = e\int_L A_\mu\, dx^\mu = -e\int_0^T A_0\, dt, \tag{16}$$

yielding a phase difference:

$$\phi_{AB} = \Delta\phi_2 - \Delta\phi_1 = -e\int_0^T A_0(t)\,dt = e\int_0^T \left(\int_{L_1}^{L_2} \mathbf{E}(\mathbf{r},t)\cdot d\mathbf{r}\right) dt, \tag{17}$$

where $\mathbf{E}(\mathbf{r},t)$ is the electric field between the two paths L_1 and L_2, and $\int_{L_1}^{L_2} \mathbf{E}(\mathbf{r},t)\cdot d\mathbf{r}$ is the gauge-invariant line integral of the electric field, representing the potential difference between the two paths at instant t, similar to the magnetic flux $\Phi(t) = \oint \mathbf{B}(t)\cdot d\mathbf{s}$ in the magnetic AB effect.

This phase accumulates continuously over $[0, T]$, mirroring the temporal buildup in the generalized magnetic AB effect driven by $\Phi(t)$. Moreover, this continuous phase accumulation offers an advantage over the magnetic AB effect. While a time-varying $\Phi(t)$ induces $\mathbf{E} = -\partial\mathbf{A}/\partial t$, perturbing trajectories, the electric AB effect maintains $\mathbf{E} = 0$ along the paths via shielding, despite $A_0(t)$'s variation inducing fields elsewhere. This eliminates trajectory shifts, making the phase's gradual accrual a pure manifestation of $A_0(t)$, offering a cleaner probe of potential-driven dynamics. In the following, however, I will mainly analyze the magnetic AB effect, since its standard form has been confirmed by experiments and also widely discussed in literature. For convenience, I will just say the AB effect or the generalized AB effect in brief.

3 Explanations in Terms of Gauge-Invariant Quantities

Having explored the nuances of the generalized AB effect, with its time-varying flux unveiling a continuous accrual of phase, we stand at a juncture of understanding. How does the phase shift emerge as the electron navigates the field-free region encircling the solenoid? This intriguing phenomenon challenges our intuitions about causality and reality, demanding an account that bridges the empirical and the theoretical. We turn now to explaining the AB effect through the lens of gauge-invariant quantities, seeking to anchor these quantum curiosities in the measurable and the invariant, a pursuit that promises both physical clarity and philosophical resonance, yet one we shall soon find fraught with its own tensions.

3.1 Why Gauge-Invariant Explanations Appear Preferable

Gauge-invariant explanations, which anchor the AB phase shift in measurable quantities like the magnetic flux Φ, the fields $\mathbf{E}$ and $\mathbf{B}$, or the velocity $\mathbf{v}$, rather than the gauge-dependent potentials $A_\mu = (-A_0, \mathbf{A})$, have long been favored in both physics and philosophy. This preference stems from their alignment with classical intuitions, empirical accessibility, and the foundational principles of gauge theories, as well as their philosophical appeal in terms of ontological parsimony. Before critiquing their adequacy, it is essential to understand why these explanations appear preferable, as they reflect deeply ingrained physical and philosophical commitments.

In the physical realm, gauge-invariant explanations resonate with a deep-seated commitment to empirical accessibility. Classical physics has taught us to trust in the tangible—field strengths $\mathbf{E}$ and $\mathbf{B}$, woven into the Lorentz force $\mathbf{F} = e(\mathbf{E} + \mathbf{v} \times \mathbf{B})$, dictate the motion of charges without reference to the shadowy potentials. In the AB effect, where $\mathbf{E} = 0$ and $\mathbf{B} = 0$ along the electron's path, the phase shift $\phi_{AB} = e\Phi$ finds its anchor in the flux $\Phi = \oint_C \mathbf{A} \cdot d\mathbf{r}$, a quantity we can grasp through the magnetic field within the solenoid. Experiments, such as those of Tonomura and colleagues in 1986, affirm this shift, lending credence to Φ as a measurable reality, a beacon of observability where $\mathbf{A}$ remains elusive. This empirical grounding aligns with a broader principle: gauge invariance, a symmetry that ensures our theories stand firm regardless of the arbitrary gauge function $\Lambda(x)$. By resting on Φ, these explanations sidestep the ambiguity of $\mathbf{A}$, which shifts with every gauge choice, offering instead a stable, invariant foundation that mirrors the elegance of Maxwell's equations and quantum electrodynamics.

Philosophically, the preference for gauge-invariant explanations deepens into questions of being and knowing. Ockham's razor, that venerable guide, urges us toward parsimony, and here it finds satisfaction. The potentials A_μ, mutable under gauge transformations, seem redundant—multiple forms yielding the same $\mathbf{E}$ and $\mathbf{B}$, yet lacking direct witness in our instruments. Why grant them ontological weight when Φ, tied to the solenoid's measurable field, suffices? This lean ontology aligns with empirical realism, a stance that demands our theories reflect the observable. Φ and $\mathbf{v}$ stand in the light of measurement, while $\mathbf{A}$ lingers in shadow, a tool of calculation rather than a pillar of reality. To the realist, this is a virtue: explanations should rest on what we can see and touch, not on constructs that shift with every mathematical whim. And in this, there is a broader coherence: gauge-invariant explanations echo the classical ontology of fields, preserving continuity with a tradition that has shaped our understanding of nature, resisting the inflation of reality with unobservable potentials.

Thus, gauge-invariant explanations beckon with a dual promise: in physics, they offer empirical grounding and theoretical symmetry; in philosophy, they deliver parsimony, realism, and a harmony with classical intuitions. However, as subsequent subsections reveal, these accounts falter under scrutiny. The nonlocal and discontinuous threads of these accounts, their failure to capture the phase's gradual unfolding, invite us to question their sufficiency—a question that will lead us, ultimately, to the potentials themselves as the deeper truth.

3.2 Gauge-Invariant Quantities in Quantum Mechanics

In order to assess the gauge-invariant explanations of the AB effect, we need to first identify the basic gauge-invariant quantities for the electron. In quantum mechanics, the Schrödinger equation governs the electron's dynamics, and its gauge invariance under transformations $A_\mu \to A'_\mu = A_\mu - \partial_\mu \Lambda$ and $\psi \to \psi' = e^{-ie\Lambda}\psi$ suggests that physical predictions should depend only on quantities unaffected by the choice of gauge function Λ. Here, I demonstrate that the basic gauge-invariant quantities for the electron are the probability density $\rho = |\psi|^2$ and the velocity $\mathbf{v} = \frac{1}{m}(\nabla S - e\mathbf{A})$, where $\psi = Re^{iS}$ is the polar form of the wave function, with R and S real-valued functions.[1]

[1] Besides these two gauge-invariant quantities, there is also a gauge-invariant energy $-(\partial_t S + eA_0)$. Since it is irrelevant to the explanations of the (magnetic) AB effect, I will not discuss it below (see, however, Gao (2025b)).

Consider the transformation of the wave function: if $\psi = Re^{iS}$, then $\psi' = e^{-ie\Lambda}\psi = Re^{i(S-e\Lambda)}$. The probability density transforms as:

$$\rho' = |\psi'|^2 = |e^{-ie\Lambda}\psi|^2 = |e^{-ie\Lambda}|^2|\psi|^2 = |\psi|^2 = \rho, \tag{18}$$

since $|e^{-ie\Lambda}| = 1$. Thus, $\rho = |\psi|^2$ is manifestly gauge-invariant, reflecting the physical requirement that the probability distribution of the electron's position remains independent of gauge choice. This invariance is local, as ρ is defined pointwise in space and time, making it a candidate for describing the AB effect without reference to the gauge-dependent potential $\mathbf{A}$.

Next, let's examine the velocity, which arises from the probability current in the continuity equation. The current density is given by:

$$\mathbf{j} = \frac{1}{2m}\left[\psi^*(\nabla - ie\mathbf{A})\psi - \psi(\nabla + ie\mathbf{A})\psi^*\right]. \tag{19}$$

Substituting $\psi = Re^{iS}$, we compute:

$$(\nabla - ie\mathbf{A})\psi = (\nabla Re^{iS} - ie\mathbf{A}Re^{iS}) = (e^{iS}\nabla R + iRe^{iS}\nabla S - ie\mathbf{A}Re^{iS}), \tag{20}$$

and its complex conjugate:

$$(\nabla + ie\mathbf{A})\psi^* = (e^{-iS}\nabla R - iRe^{-iS}\nabla S + ie\mathbf{A}Re^{-iS}). \tag{21}$$

The current becomes:

$$\mathbf{j} = \frac{R^2}{m}(\nabla S - e\mathbf{A}), \tag{22}$$

since the imaginary terms cancel. Defining the velocity as $\mathbf{v} = \frac{\mathbf{j}}{\rho}$, with $\rho = R^2$, we obtain:

$$\mathbf{v} = \frac{1}{m}(\nabla S - e\mathbf{A}). \tag{23}$$

Under a gauge transformation, $\mathbf{A} \to \mathbf{A}' = \mathbf{A} - \nabla\Lambda$ and $S \to S' = S - e\Lambda$, so:

$$\nabla S' = \nabla(S - e\Lambda) = \nabla S - e\nabla\Lambda, \tag{24}$$

and thus:

$$\mathbf{v}' = \frac{1}{m}(\nabla S' - e\mathbf{A}') = \frac{1}{m}[(\nabla S - e\nabla\Lambda) - e(\mathbf{A} - \nabla\Lambda)] = \frac{1}{m}(\nabla S - e\mathbf{A}) = \mathbf{v}. \tag{25}$$

The velocity $\mathbf{v}$ is therefore gauge-invariant, as the transformation of ∇S cancels that of $\mathbf{A}$. Like ρ, $\mathbf{v}$ is a local quantity, defined at each spacetime point, and together they fully characterize the electron's dynamical state in a gauge-independent manner.

These quantities—ρ and $\mathbf{v}$—are two basic gauge-invariant descriptors of the quantum state, analogous to classical observables like density and velocity in hydrodynamics. In the context of the AB effect, gauge-invariant explanations rely on them to attribute the phase shift $\phi_{AB} = e\Phi$ to measurable properties, such as the flux Φ influencing $\mathbf{v}$ at interference. However, as subsequent subsections reveal, their inability to account for the phase's continuous accumulation—due to their insensitivity to $\mathbf{A}$'s local influence before overlap—exposes the limitations of this approach, setting the stage for the critique in Section 3.3 and beyond.

3.3 Dynamics Before and After Overlapping

Consider the standard AB setup: two electron beams encircle a solenoid with constant magnetic flux Φ, recombining to interfere. Before overlap, each beam travels in a simply connected, field-free region ($\mathbf{B} = 0$), where a gauge choice $\mathbf{A} = 0$ is possible. In this gauge, the Schrödinger equation reduces to the free form, and the solutions ψ_1 and ψ_2 for each beam match those of a free electron, implying ρ and $\mathbf{v}$ are independent of Φ. This holds because, in each path, the gauge transformation adjusts the phase locally, leaving gauge-invariant properties unchanged.

However, after the beams overlap, forming a closed loop C around the solenoid, $\mathbf{A} = 0$ cannot be chosen globally due to the nonzero flux $\Phi = \oint_C \mathbf{A} \cdot d\mathbf{r}$. The interference pattern shifts by:

$$\phi_{AB} = e \oint_C \mathbf{A} \cdot d\mathbf{r} = e\Phi, \tag{26}$$

and the velocity satisfies

$$\oint_C \mathbf{v} \cdot d\mathbf{r} = \oint_C \frac{1}{m}(\nabla S - e\mathbf{A}) \cdot d\mathbf{r} = -e\Phi \tag{27}$$

reflecting Φ's influence. Consequently, $\mathbf{v}$ and ρ (via the continuity equation $\partial_t \rho + \nabla \cdot (\rho \mathbf{v}) = 0$) abruptly depend on Φ at overlap, despite being Φ-independent beforehand. This transition is instantaneous and discontinuous, suggesting a sudden nonlocal influence from the solenoid's magnetic field or current.

3.4 Critique of Gauge-Invariant Explanations

The gauge-invariant approach to the AB effect, which privileges measurable quantities such as the magnetic flux Φ or the velocity field $\mathbf{v}$ over the gauge-dependent vector potential $\mathbf{A}$, encounters a series of philosophical and physical objections that collectively undermine its explanatory legitimacy. This perspective, rooted in a classical sensibility that assigns ontological primacy to field strengths or their proxies, seeks to preserve a reality defined by what is directly observable, eschewing the apparent arbitrariness of potentials. Yet, when subjected to critical scrutiny, this stance reveals profound shortcomings that clash with the dynamic nature of quantum phenomena as revealed by the AB effect. The following paragraphs delineate these objections, exposing the nonlocality, discontinuity, and incompleteness inherent in gauge-invariant explanations, and thereby necessitating a reevaluation of the physical ontology they purport to uphold.

One primary objection lies in the nonlocality implicit in attributing the AB phase shift to Φ, a global property confined within the solenoid. This approach posits that the phase ϕ_{AB} emerges instantaneously at the point of interference, reflecting an action at a distance on the electron despite its confinement to a field-free region—a proposition that strains the causal architecture of special relativity, which insists that physical effects propagate no faster than the speed of light. Such an unmediated action across space suggests a reality where distant entities can affect one another without a local intermediary, a notion that sits uneasily with the principle of locality. This nonlocality not only violates relativistic constraints but also introduces a metaphysical perplexity: how can a physical state, ostensibly grounded in observable quantities, depend on a distant configuration without a discernible causal link? The gauge-invariant insistence on Φ as the sole arbiter of the phase shift thus falters, failing to reconcile the effect's spatial dynamics with a coherent account of physical causation.

A further objection arises from the discontinuity exhibited by gauge-invariant quantities such as $\rho = |\psi|^2$ and $\mathbf{v} = \frac{1}{m}(\nabla S - e\mathbf{A})$ in their temporal evolution. Prior to the beams' overlap, these quantities reflect the undisturbed state of a free electron, showing no trace of Φ's influence, as established in Section 3.3; yet, at the moment of interference, they abruptly adjust to incorporate the flux, satisfying relations like $\oint_C \mathbf{v} \cdot d\mathbf{r} = -e\Phi$. This sudden shift stands in stark contrast to the expectation in quantum mechanics that physical states evolve smoothly unless perturbed by local interactions—a principle of continuity that underpins the theory's predictive coherence. Philosophically, this discontinuity poses a challenge to the gauge-invariant ontology: if reality is to be captured by measurable properties alone, why do these properties undergo an inexplicable leap, devoid of a gradual process to bridge the pre- and post-overlapping states? The absence of a dynamic mechanism renders the explanation ontologically brittle, unable to account for the temporal becoming that quantum phenomena demand.

Perhaps the most damning objection is the incompleteness of gauge-invariant quantities in elucidating the phase's origin during the electron's transit. Before interference, ρ and $\mathbf{v}$ remain indifferent to Φ, offering no insight into how the phase accumulates as the electron navigates its path; only at overlap does the shift ϕ_{AB} manifest, yet the process by which it arises remains opaque within this framework. In contrast, quantum mechanics attributes the phase $\phi_{AB} = e\oint_C \mathbf{A} \cdot d\mathbf{r}$ to the vector potential $\mathbf{A}$, a dependence that traces its continuous generation across the trajectory—an account the gauge-invariant perspective cannot replicate without invoking

A itself. This limitation exposes an epistemological shortfall: by restricting physical reality to what is gauge-invariant, the approach overlooks the underlying dynamics that quantum theory reveals, rendering it incapable of fully explaining the AB effect's causal structure. The insistence on Φ or $\mathbf{v}$ as sufficient descriptors thus betrays a failure to engage with the effect's temporal and spatial unfolding, necessitating a turn toward the potentials' explanatory power.

To sum up, these objections—nonlocality, discontinuity, and incompleteness—collectively dismantle the gauge-invariant explanation of the AB effect, revealing its inability to provide a philosophically robust account of the phase shift's generation. Far from preserving a coherent ontology, this approach sacrifices the locality, continuity, and completeness demanded by quantum mechanics, compelling a reconsideration of the vector potential's role as more than a mathematical convenience. This critique sets the stage for a potential-centric ontology, pursued in subsequent sections, that better aligns with the AB effect's empirical and theoretical reality.

3.5 Madelung Hydrodynamics as a Paradigm

The critique of gauge-invariant explanations in Section 3.4 exposed their inherent flaws in addressing the AB effect. These approaches, which elevate measurable quantities like the magnetic flux Φ over the vector potential $\mathbf{A}$, struggle to provide a dynamic explanation of the phase's genesis, highlighting a systemic limitation in gauge-invariant paradigms. A typical example of such an approach is the Madelung hydrodynamic formulation, which serves as a paradigmatic case inheriting the same trio of problems identified above.

Proposed by Erwin Madelung in 1927, this framework offers a gauge-invariant reinterpretation of quantum mechanics, recasting the Schrödinger equation in terms of fluid-like quantities—probability density ρ and velocity $\mathbf{v}$—independent of electromagnetic potentials. At first sight, this approach appears primed to explain the AB effect locally and continuously, avoiding the gauge dependence of $\mathbf{A}$. However, a closer examination reveals profound limitations: the Madelung equations, supplemented by a nonlocal quantization condition, fail to provide a coherent, physically meaningful account of the AB phase shift. This section derives the Madelung framework, exposes its inequivalence to the Schrödinger equation, and critiques its inability to explain the AB effect without invoking nonlocality and discontinuity (for a more detailed analysis see Gao, 2025c).

3.5.1 Derivation of the Madelung Equations

Starting from the Schrödinger equation (Equation (1) in Section 2), let $\psi = Re^{iS}$, where $R = \sqrt{\rho}$ and S is the phase. Substituting into Equation (1) and separating real and imaginary parts (assuming $\psi \neq 0$) yields two coupled equations:

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{v}) = 0, \tag{28}$$

where $\mathbf{v} = \frac{1}{m}(\nabla S - e\mathbf{A})$ is the velocity field, and

$$m\frac{\partial \mathbf{v}}{\partial t} = e(\mathbf{E} + \mathbf{v} \times \mathbf{B}) - m(\mathbf{v} \cdot \nabla)\mathbf{v} - \nabla U, \tag{29}$$

where $U = -\frac{1}{2m}\frac{\nabla^2 R}{R}$ is the quantum potential, $\mathbf{E} = -\nabla A_0 - \partial \mathbf{A}/\partial t$, and $\mathbf{B} = \nabla \times \mathbf{A}$. Equation (28) is the continuity equation, ensuring probability conservation, while Equation (29) resembles a hydrodynamic momentum equation, with U introducing quantum effects.

These Madelung equations involve only gauge-invariant quantities (ρ, $\mathbf{v}$, $\mathbf{E}$, $\mathbf{B}$), suggesting a potential-free description of the AB effect. Yet, their equivalence to the Schrödinger equation hinges on subtle constraints.

3.5.2 Inequivalence to the Schrödinger Equation

Reconstructing the Schrödinger equation from Equations (28) and (29) requires $\mathbf{v} = \nabla S/m$ (for $\mathbf{A} = 0$) and integration of Equation (29) to recover S. However, this fails globally: S, as the wave function's phase, is multi-valued in multiply connected spaces (e.g., $S = l\theta$ for angular momentum states, where θ is azimuthal angle and l is an integer). Thus, $\mathbf{v}$ cannot always be the gradient of a single-valued S. Without additional constraints, $\psi = Re^{iS}$ may be multi-valued, violating the Schrödinger equation's requirement of single-valuedness.

To align the Madelung equations with quantum mechanics, a quantization condition is imposed:

$$m \oint_C \mathbf{v} \cdot d\mathbf{r} = 2\pi n - e\Phi, \tag{30}$$

where C is a closed loop, n is an integer, and $\Phi = \oint_C \mathbf{A} \cdot d\mathbf{r}$ is the enclosed magnetic flux. With $\mathbf{v} = \frac{1}{m}(\nabla S - e\mathbf{A})$, this ensures ψ's single-valuedness by fixing $\Delta S = S(\theta + 2\pi) - S(\theta) = 2\pi n$. While this restores formal equivalence, it introduces a nonlocal constraint, as $\mathbf{v}$ along C depends on Φ inside the solenoid.

3.5.3 Nonlocality of the Quantization Condition

The condition in Equation (30) is global, not local: it integrates $\mathbf{v}$ over a closed path, linking it to Φ in a distant region without a mediating field. This resembles the Coulomb gauge's nonlocal potential dependence (Wallace, 2024), undermining claims of a local explanation. Physically, it lacks a clear mechanism—unlike the Schrödinger equation, where $\mathbf{A}$ locally drives the phase, the quantization condition imposes a mathematical fix without intrinsic justification (Wallstrom, 1994). Even proponents like Takabayasi (1952) acknowledged its ad hoc nature, highlighting its explanatory deficit.

3.5.4 Failure to Explain the AB Effect

Applying Madelung hydrodynamics to the magnetic AB effect reveals its shortcomings. Before the beams overlap, the field-free region ($\mathbf{E} = \mathbf{B} = 0$) reduces Equations (28) and (29) to free-particle forms, yielding ρ and $\mathbf{v}$ independent of Φ. The quantization condition (Equation (30)) cannot apply, as no closed loop exists. At overlap, it abruptly holds, shifting $\mathbf{v}$ and ρ discontinuously to reflect Φ. This mirrors the gauge-invariant critique in Section 3.4: the phase shift emerges instantaneously, not continuously, implying nonlocal action from the solenoid.

Why this sudden influence? The Madelung framework offers no physical process—the condition is a mathematical artifact, not a dynamical explanation. Unlike the Schrödinger equation's potential-driven phase accumulation, Madelung hydrodynamics leaves the AB effect's genesis opaque, relying on an unmotivated global constraint.

To sum up, Madelung hydrodynamics, despite its gauge-invariant allure, fails as a paradigm for the AB effect. Its quantization condition introduces nonlocality and discontinuity, echoing the universal flaws of gauge-invariant approaches, while its inability to dynamically account for the phase shift underscores its incompleteness. This analysis reinforces the need for electromagnetic potentials, explored in later sections, to provide a local, continuous explanation.

3.6 Healey's Loop Ontology and Its Objections

The persistent failure of gauge-invariant explanations to account for the continuous accumulation of the AB phase, as critiqued in Sections 3.3 through 3.5, motivates exploration of alternative frameworks that might salvage a gauge-invariant ontology. Healey's loop ontology, articulated by Healey (2007), proposes one such approach by positing that the physical reality of the AB effect resides in the holonomies—phase factors $e^{ie\oint_C \mathbf{A}\cdot d\mathbf{r}}$—associated with closed loops in spacetime. This topological perspective seeks to eliminate dependence on the gauge-dependent vector potential $\mathbf{A}$, grounding the phase shift $\phi_{AB} = e\Phi$ in a property invariant under gauge transformations $A_\mu \to A_\mu - \partial_\mu \Lambda$. By assigning ontological primacy to these loop-based quantities, Healey aligns with the philosophical commitment to observable, gauge-invariant entities as the bearers of physical significance. While this perspective might appear to circumvent the conceptual difficulties plaguing the Madelung hydrodynamics, a rigorous analysis reveals that Healey's loop ontology succumbs to analogous deficiencies, failing to provide a dynamically coherent explanation of the AB effect's phase generation.

3.6.1 Dynamical Inadequacy of the Loop Ontology

First, Healey's loop ontology lacks a dynamical law to elucidate how the AB phase accumulates along the electron's trajectory. By tying reality to holonomies defined over closed loops, the framework implies that the phase shift ϕ_{AB} emerges only upon the beams' recombination at interference, with no mechanism specified for its gradual accrual during transit through the field-free region. Before overlap, the separate paths L_1 and L_2 do not form a closed loop $C = L_1 - L_2$, rendering the holonomy undefined and the potential's influence ($\mathbf{A}$) ostensibly inert in this ontology. This absence of a temporal process mirrors the discontinuity critiqued in prior gauge-invariant accounts (Section 3.3), clashing with the smooth dynamics demanded by the Schrödinger equation. Philosophically, this omission raises a profound question: how can an ontology claim explanatory power if it fails to illuminate the becoming of the very phenomenon it seeks to describe? The lack of a dynamical account renders the loop ontology a static abstraction, insufficient for capturing the AB effect's physical reality as a process unfolding in spacetime.

3.6.2 Structural Redundancy in the Loop Framework

Second, the loop ontology introduces structural redundancy that compromises its claim to ontological primacy. Assigning phase factors to loops presupposes a consistent underlying structure: for a large loop enclosing the solenoid, the holonomy $e^{ie\oint_C \mathbf{A}\cdot d\mathbf{r}} = e^{ie\Phi}$ must equal the product of holonomies over smaller sub-loops, a coherence that implicitly relies on the potential $\mathbf{A}$ or its field $\mathbf{B} = \nabla \times \mathbf{A}$. In simply connected regions outside the solenoid, where $\mathbf{B} = 0$, the holonomies could be trivially unity if $\mathbf{A}$ were gauged away, yet the nonzero flux Φ necessitates $\mathbf{A} \neq 0$ globally (Section 2.1). This dependence suggests that the loop values are not fundamental but derivative, encoding the same information as $\mathbf{A}$ in a more convoluted form (Wallace, 2014). Rather than supplanting the potential, Healey's framework rephrases its effects topologically, inflating ontological complexity without resolving the causal question of phase generation. This redundancy echoes the Madelung approach's reliance on a quantization condition (Section 3.5.3), weakening its status as a standalone explanation.

3.6.3 Nonlocal Interdependence and Causal Concerns

Third, the loop ontology entails a non-separable interdependence that introduces a subtle yet pervasive nonlocality, undermining its compatibility with relativistic causality. Consider two simply connected regions X and Y outside the solenoid, whose union $X \cup Y$ encloses the flux and becomes multiply connected. Specifying holonomies for all loops within X and Y individually—where $\mathbf{B} = 0$ and local gauge choices might nullify $\mathbf{A}$—does not determine the holonomy for a loop in $X \cup Y$ encircling Φ, which depends on the distant flux without a local mediator (Wallace, 2014). This non-separability implies that the solenoid's configuration instantaneously influences the phase across disjoint regions, echoing the nonlocal action critiqued in Section 3.4.

In conclusion, Healey's framework of loop ontology, despite its topological elegance, shares the Madelung approach's nonlocal and discontinuous traits, failing to resolve the AB effect's dynamical origin while adding ontological complexity. This re-affirms the need for a potential-centric perspective, pursued in subsequent sections, to address the AB effect's reality with the depth and locality it demands.

4 A No-Go Result for Gauge-Invariant Explanations

The AB effect poses a profound challenge to our understanding of quantum mechanics, particularly in how we interpret the physical significance of electromagnetic potentials versus gauge-invariant quantities. Proponents of gauge-invariant explanations contend that the AB phase shift arises instantaneously at the point of interference, a position critiqued in Section 3 for its reliance on nonlocal and discontinuous mechanisms. This section advances a more decisive argument grounded in the generalized AB effect and supported by proposed experiments. The generalized AB effect demonstrates that the phase accumulates continuously as the electron moves along its path, rather than emerging instantaneously at interference—a finding consistent with quantum mechanics' predictions. Since gauge-invariant explanations presuppose an instantaneous phase shift, they are not merely inadequate but *excluded* by this evidence, compelling a reevaluation of the ontology underlying quantum phenomena.

4.1 An Incompatibility Proof

This section constructs a formal no-go theorem, proving that explanations relying solely on gauge-invariant quantities—such as the magnetic flux Φ or derived fields—are untenable. The proof hinges on a contradiction between two propositions: (1) gauge-invariant accounts necessitate an instantaneous phase shift at interference, and (2) the generalized AB effect demonstrates continuous phase accumulation along the electron's path. The arguments supporting these propositions have been presented earlier—specifically, in Sections 3.3 and 2.2, respectively—and are restated here for clarity and rigor. I substantiate each proposition, derive their contradiction, and address potential objections to cement the exclusion of gauge-invariant paradigms.

Proposition 1: Gauge-Invariant Explanations Require Instantaneous Phase Shift at Interference

Gauge-invariant explanations assert that the AB phase shift ϕ_{AB} arises solely from measurable, gauge-independent quantities, such as the magnetic flux $\Phi = \oint_C \mathbf{B} \cdot d\mathbf{s}$, and manifests only when the electron beams recombine at interference. The argument was given in Section 3.3, and I restate it formally here.

Consider the standard AB setup: a coherent electron beam splits at $t = 0$ into two paths, L_1 and L_2, encircling a solenoid with constant flux Φ, recombining at $t = T$. Before overlap, each beam travels in a simply connected, field-free region ($\mathbf{B} = 0$) outside the solenoid. In such a region, a gauge choice $\mathbf{A} = 0$ is permissible, as the magnetic field's vanishing curl ($\nabla \times \mathbf{A} = 0$) allows the vector potential to be gauged away via:

$$\mathbf{A}' = \mathbf{A} - \nabla\Lambda, \quad \Lambda(\mathbf{r}) = \int_{\mathbf{r}_0}^{\mathbf{r}} \mathbf{A}(\mathbf{r}') \cdot d\mathbf{r}', \tag{31}$$

where $\mathbf{r}_0$ is a reference point, and $\nabla \times \mathbf{A} = 0$ ensures Λ is well-defined in the simply connected domain of each path. In this gauge, with $A_0 = 0$ outside, the Schrödinger equation (1) reduces to the free-particle form:

$$i\frac{\partial\psi}{\partial t} = -\frac{1}{2m}\nabla^2\psi, \tag{32}$$

since $\mathbf{A} = 0$. The solutions ψ_1 and ψ_2 for each beam match those of a free electron (e.g., plane waves $\psi_i = e^{i(\mathbf{k}\cdot\mathbf{r}-\omega t)}$), devoid of any phase dependence on Φ, as $\mathbf{B} = 0$ locally and Φ is confined to the solenoid's interior, inaccessible to the beams.

At interference ($t = T$), the beams overlap, forming a closed loop $C = L_1 - L_2$ around the solenoid, a multiply connected region where $\mathbf{A} = 0$ cannot be globally maintained due to:

$$\oint_C \mathbf{A} \cdot d\mathbf{r} = \Phi \neq 0. \tag{33}$$

The interference pattern shifts by:

$$\phi_{AB} = e\oint_C \mathbf{A} \cdot d\mathbf{r} = e\Phi, \tag{34}$$

reflecting Φ's influence. As argued in Section 3.3, gauge-invariant accounts interpret this as an instantaneous effect: before overlap, ψ_1 and ψ_2 evolve as free particles ($\rho = |\psi|^2$ and $\mathbf{v} = \frac{\nabla S}{m}$, with no Φ-dependence), and only at $t = T$ does Φ—a nonlocal, measurable quantity—impose the phase shift. This requires the phase to emerge discontinuously at interference, as no local mechanism (e.g., $\mathbf{B}$) alters the beams' evolution beforehand, and Φ's effect is tied to the final recombination event.

Proposition 2: Generalized AB Effect Demonstrates Continuous Phase Accumulation

In the standard magnetic AB effect, a constant flux Φ produces a phase shift $\phi_{AB} = e\Phi$ (Section 2.1), which gauge-invariant accounts attribute to an immediate effect of Φ at interference. The generalized AB effect, where the flux $\Phi(t)$ varies with time, contradicts this instantaneous picture by showing the phase accumulates continuously as the electron traverses its path. This argument was developed in Section 2.2, and we restate it with formal detail here.

Consider the setup: an electron beam splits at $t = 0$, follows paths L_1 and L_2 around a solenoid with time-dependent flux $\Phi(t)$, and recombines at $t = T$. In a gauge with $A_0 = 0$,

$$\mathbf{A}(\mathbf{r}, t) = \frac{\Phi(t)}{2\pi r}\hat{\boldsymbol{\theta}}, \tag{35}$$

ensuring $\oint_C \mathbf{A} \cdot d\mathbf{r} = \Phi(t)$. For circular paths of radius R, with $d\mathbf{r} = R\omega_i(t)\hat{\boldsymbol{\theta}}dt$ and $\omega_1(t) + \omega_2(t) = \frac{2\pi}{T}$ (as derived in Section 2.2), the phase difference is:

$$\phi_{AB} = e\left(\int_{L_2} \mathbf{A} \cdot d\mathbf{r} - \int_{L_1} \mathbf{A} \cdot d\mathbf{r}\right) = \frac{1}{T}\int_0^T e\Phi(t)\, dt. \tag{36}$$

To test this, define:

$$\Phi(t) = \begin{cases} \Phi & \text{for } 0 \le t \le T/2, \\ 0 & \text{for } T/2 < t \le T, \end{cases} \tag{37}$$

yielding:

$$\phi_{AB} = \frac{1}{T}\int_0^{T/2} e\Phi\, dt = \frac{e\Phi}{2}. \tag{38}$$

This phase, observable at interference, depends on $\Phi(t)$'s profile over $0 \le t \le T$, not just $\Phi(T) = 0$.

Contradiction and No-Go Theorem

The contradiction is now clear. Proposition 1 demands that ϕ_{AB} emerges instantaneously at $t = T$, determined by $\Phi(T)$:

$$\phi_{AB} = e\Phi(T) = 0, \tag{39}$$

since $\Phi(T) = 0$. Proposition 2, validated by Equation (38), yields:

$$\phi_{AB} = \frac{e\Phi}{2} \neq 0, \tag{40}$$

reflecting the continuous integration of $\Phi(t)$. These cannot coexist: if the phase were instantaneous, it would vanish when $\Phi(T) = 0$, yet the generalized AB effect's non-zero result proves it accumulates over the path. Formally, define the gauge-invariant hypothesis H_{GI}: "The phase shift ϕ_{AB} is a function of Φ at interference, $\phi_{AB} = f(\Phi(T))$, with no prior accumulation." The generalized AB effect falsifies H_{GI}, as $\phi_{AB} = \frac{1}{T}\int_0^T e\Phi(t)\,dt \neq f(\Phi(T))$, establishing the no-go theorem: gauge-invariant explanations are impossible for the AB effect.

4.2 Objections and Rebuttals

Two objections might challenge this proof, but both fail under scrutiny. First, proponents like Healey (2007) argue that $\phi_{AB} = e\oint_C \mathbf{A}\cdot d\mathbf{r}$ is a gauge-invariant holonomy, a loop property manifesting at interference without requiring continuous accumulation. In the static case, this holds ($\phi_{AB} = e\Phi$), but for $\Phi(t)$, the holonomy:

$$e^{ie\oint_C \mathbf{A}\cdot d\mathbf{r}} = e^{i\frac{1}{T}\int_0^T e\Phi(t)\,dt}, \tag{41}$$

varies with the time integral, not $\Phi(T)$. The phase's dependence on $\Phi(t)$'s history—e.g., $\frac{e\Phi}{2}$ versus 0—cannot be a static loop property evaluated at $t = T$. The objection thus misapplies a static concept to a time-varying context, failing to account for the accumulation evidenced by Equation (38).

Next, one might posit that $\Phi(t)$ nonlocally influences ψ_1 and ψ_2 before overlap, bypassing the $\mathbf{A} = 0$ gauge. However, in a simply connected region ($\mathbf{B} = 0$), the gauge transformation (Equation 31) eliminates $\mathbf{A}$, and no local field couples Φ to the beams. Moreover, nonlocality violates special relativity's causality (effects propagate at most at c), and quantum mechanics offers no such mechanism—Φ's effect requires the closed loop at interference, yet its timing contradicts the continuous buildup. This objection thus introduces an ad hoc, unphysical assumption, refuted by the gauge invariance of the Schrödinger equation.

In conclusion, the no-go theorem stands: gauge-invariant explanations, which predict an instantaneous phase shift at interference, are incompatible with the generalized AB effect's continuous phase accumulation. This contradiction excludes such accounts, compelling a shift to electromagnetic potentials as the fundamental reality, a foundation explored in Section 6.

4.3 Experimental Proposals

In order to empirically validate the exclusion of gauge-invariant explanations and affirm the primacy of electromagnetic potentials, we propose two experimental designs leveraging modern interferometric precision. These experiments test the phase shift's dependence on the time-varying magnetic flux $\Phi(t)$'s history, contrasting it with the gauge-invariant expectation tied to $\Phi(T)$ at interference. Below, we detail each setup, derive expected outcomes, address experimental challenges, providing a concrete path to resolve this ontological dispute.

Experiment 1: Time-Varying Flux with Mid-Transit Shut-Off

This experiment directly probes the generalized AB effect by manipulating $\Phi(t)$ during the electron's transit, testing whether the phase reflects the flux's temporal profile or its value at interference.

Setup: A coherent electron beam, emitted from a source at $t = 0$, is split by a biprism into two paths, L_1 and L_2, encircling an infinitely long solenoid of radius R_s (e.g., $R_s \approx 1\,\mu\text{m}$), recombining at a detector after transit time T (e.g., $T \approx 10^{-8}\,\text{s}$, based on electron velocities $\sim 10^7\,\text{m/s}$ over a $\sim 0.1\,\text{m}$ path). The solenoid, shielded to confine $\mathbf{B}$ internally, generates a flux $\Phi(t)$ controlled by a precise current source. Initially, $\Phi(t) = \Phi_0$ (e.g., $\Phi_0 = 10^{-15}\,\text{Wb}$) from $t = 0$ to $t = T/2$, then switches to $\Phi(t) = 0$ from $t = T/2$ to T, with a transition duration $\Delta t \approx 10^{-9}\,\text{s}$.

Theoretical Predictions: Quantum mechanics predicts the phase shift via Equation (38):

$$\phi_{AB} = \frac{1}{T}\int_0^{T/2} e\Phi_0\,dt = \frac{e\Phi_0}{2}, \tag{42}$$

since $\Phi(t) = 0$ for $T/2 < t \leq T$. This reflects the flux's influence over the first half of the transit, accumulated continuously via $\mathbf{A}$. Gauge-invariant explanations, assuming the phase depends on $\Phi(T)$ at interference ($t = T$), predict:

$$\phi_{AB} = e\Phi(T) = 0, \tag{43}$$

as $\Phi(T) = 0$. The interference pattern's fringe shift, proportional to ϕ_{AB}, will distinguish these outcomes: a shift of $\frac{e\Phi_0}{2h}\lambda$ (where λ is the electron wavelength) versus no shift.

Experimental Considerations: The solenoid's current must switch from $I_0 = \frac{\Phi_0}{\mu_0 \pi R_s^2}$ to 0 at $t = T/2$, synchronized with the electron's transit. Timing precision ($\pm 10^{-10}\,\text{s}$) is critical, achievable with pulsed current sources and fast-switching transistors. Electron coherence requires a low-temperature environment and minimal path-length differences, standard in electron interferometry (Tonomura et al., 1986). The transition induces a transient $\mathbf{E} = -\frac{\partial \mathbf{A}}{\partial t}$, but Section 2.2 shows its effect cancels in ϕ_{AB}, ensuring the phase depends only on $\Phi(t)$'s integral.

Experiment 2: Relativistic Propagation of Phase Shift with Flux Shut-Off

This experiment exploits the finite propagation speed of the influence of the current or flux inside the solenoid on the electron to test whether the phase shift reflects this retarded influence, further challenging gauge-invariant nonlocality.

Setup: Similar to Experiment 1, an electron beam splits at $t = 0$, encircles a solenoid, and recombines at $t = T$ (e.g., $T = 10^{-8}\,\text{s}$) at a distance $d = 0.3\,\text{m}$ from the solenoid. The solenoid maintains $\Phi(t) = \Phi_0$ until $t = T - \delta$, where $\delta = d/c \approx 10^{-9}\,\text{s}$, then switches to $\Phi(t) = 0$ via a current shut-off at $t = T - \delta$.

Theoretical Predictions: If the flux inside the solenoid influences the electron at finite propagation speed of light (e.g via $\mathbf{A}$ in the Lorenz gauge), obeying special relativity, then the effect of its shut-off at $t = T - \delta$ will not reach the electron beams before $t = T$, as the distance d ensures a light-travel delay. Thus the AB phase shift will be

$$\phi_{AB} = \frac{1}{T}\int_0^T e\Phi(t)\,dt = e\Phi_0, \tag{44}$$

as in the standard AB effect. In fact, when considering the relativistic retardation effect, the generalized AB phase shift should be

$$\phi_{AB} = \frac{1}{T}\int_0^T e\Phi(t - R/c)\,dt, \tag{45}$$

where R is the radius of the electron's circular paths. In contrast, gauge-invariant accounts, assuming instantaneous dependence on $\Phi(T) = 0$ at interference, predict:

$$\phi_{AB} = e\Phi(T) = 0. \tag{46}$$

The fringe shift ($\frac{e\Phi_0}{h}\lambda$ vs. 0) tests whether $\Phi(t)$'s retarded propagation, not $\Phi(T)$, determines ϕ_{AB}.

Experimental Considerations: Synchronization between the shut-off (at $t = T - 10^{-9}\,\text{s}$) and electron emission must be precise ($\pm 10^{-10}\,\text{s}$). The shut-off's finite speed ($\Delta t \approx 10^{-10}\,\text{s}$) introduces a small error in Equation (38), reducible to $\sim \frac{\Delta t}{T}\Phi_0 \approx 10^{-2}\Phi_0$, negligible for detection.

Both experiments leverage $\Phi(t)$'s controllability to probe the phase's genesis. Experiment 1 tests temporal accumulation within a compact setup, ideal for laboratory precision, while Experiment 2 probes relativistic causality, requiring higher timing precision but offering deeper ontological insight. Success in either—showing $\phi_{AB} \neq e\Phi(T)$—would empirically falsify gauge-invariant instantaneity. Challenges like coherence and timing are surmountable with current technology, positioning these proposals as practical tests of the AB effect's foundational implications.

4.4 Conclusion

To summarize, the generalized AB effect, buttressed by its experimental implications, lays bare the inadequacy of gauge-invariant explanations with unassailable clarity. The evidence that the phase accumulates continuously as the electron traverses its path, rather than emerging instantaneously at interference—a phenomenon quantum mechanics consistently predicts—excludes gauge-invariant interpretations, which hinge on an untenable instantaneity. Quantities like Φ, ρ, and $\mathbf{v}$, bound to static or abrupt manifestations, cannot account for this dynamic process, which finds its root in $\mathbf{A}$'s local influence (Section 2.2). This exclusion compels a philosophical shift, directing us to reconsider electromagnetic potentials as ontologically fundamental—a path pursued in Section 6.

5 Is Quantum Electrodynamics Relevant?

The analysis thus far has rested on a quantum mechanical (QM) description of the AB effect, treating the electron as a quantum particle interacting with classical electromagnetic potentials. This approach, while fruitful, prompts a critical question: does a fully quantized treatment in quantum electrodynamics (QED)—where both the electron and electromagnetic field are quantum entities—alter the conclusion that gauge-invariant explanations are inadequate? QED, as the quantum field theory unifying QM and special relativity, offers a more complete framework for electromagnetic interactions, potentially reframing the AB effect's ontology. This section argues that QED reinforces, rather than undermines, the critique of gauge-invariant explanations established in Sections 3–4, demonstrating that the phase shift's continuous accumulation remains tied to potentials, not gauge-invariant quantities, even in a quantized context.

5.1 Correspondence Between QM and QED

A foundational observation underpins this inquiry: QM serves as a robust approximation to QED in regimes relevant to the AB effect. The correspondence principle ensures that QM's predictions—e.g., the phase shift $\phi_{AB} = e \oint_C \mathbf{A} \cdot d\mathbf{r}$ in the magnetic AB effect—closely align with QED's, with quantum corrections typically small. For instance, vacuum polarization due to the solenoid's magnetic field introduces an exponentially suppressed effect beyond the electron's Compton wavelength ($\lambda_C \approx 2.43 \times 10^{-12}\,\mathrm{m}$), negligible in typical AB setups where distances exceed this scale (Serebryanyi, 1985; Gornicki, 1990). In QM, gauge-invariant quantities like $\rho = |\psi|^2$ and $\mathbf{v} = \frac{1}{m}(\nabla S - e\mathbf{A})$ fail to account for the phase's continuous emergence (Sections 3-4); their QED counterparts—e.g., expectation values of charge density or current—inherit this limitation, as their dynamics remain tied to the classical $\mathbf{A}$'s influence, minimally perturbed by quantization. Thus, QED's corrections do not introduce new gauge-invariant mechanisms sufficient to explain the AB phase, preserving QM's insight that potentials drive the effect.

5.2 Gauge Invariance in a QED Framework

Could QED offer a gauge-invariant explanation unavailable in QM, perhaps via quantized field interactions? Marletto and Vedral (2020) propose that the AB phase arises locally from the interaction energy between a charged particle and a current source, mediated by photon exchange—a quantity often considered gauge-invariant. They argue that this energy shift, measurable even for non-closed paths, obviates the need for potentials. However, this claim falters under scrutiny. Hayashi (2023) demonstrates that, in QED, the interaction energy is gauge-dependent, varying with the choice of gauge (e.g., Coulomb vs. axial gauge) due to the photon propagator's form. For a non-closed path, the phase shift $\Delta\phi = e \int_L A_\mu dx^\mu$ remains gauge-variant, as Wakamatsu (2024) confirms using a path-integral approach with an effective Lagrangian. Saldanha (2024), correcting an earlier error (Saldanha, 2021), reaffirms this: only for closed loops does the phase become gauge-invariant, as $\phi_{AB} = e \oint_C \mathbf{A} \cdot d\mathbf{r}$. Thus, QED does not yield a local, gauge-invariant phase for the electron's transit, reinforcing that $\mathbf{A}$'s role persists, consistent with the continuous accumulation observed in the generalized AB effect (Section 2.2).

5.3 Vaidman's Alternative Proposal

Vaidman (2012) offers a distinct perspective, suggesting that the AB phase emerges from local interactions between the electron's field and the solenoid's current, eschewing potentials. He posits that the electron exerts an electric force on the solenoid's charges, inducing a phase shift in their joint wave function. Yet, Pearle and Rizzi (2017) show this to be equivalent to the standard QM account: the net phase, including the electron's standard shift (ϕ_{AB}), Vaidman's contribution, and an opposing interaction term, sums to ϕ_{AB}, leaving the potential-based explanation intact (see also Aharonov and Bohm, 1961). Aharonov et al. (2016) further point out that Vaidman's account does not apply in sourceless AB scenarios (e.g., magnetic flux without current), where no interacting charges exist to mediate the effect. In QED, where fields are operator-valued, the photon field's potential A_μ remains essential to the Hamiltonian, coupling to the electron's charge regardless of source dynamics. Vaidman's proposal, while intriguing, does not supplant $\mathbf{A}$'s primacy, aligning with QED's reinforcement of QM's potential-driven ontology.

5.4 Implications for Gauge-Invariant Explanations

The continuity between QM and QED, coupled with the failure of QED-specific gauge-invariant proposals, solidifies the critique of Sections 3–4. In QED, the AB phase—whether $\phi_{AB} = e\Phi$ or its generalized form $\phi_{AB} = \frac{1}{T} \int_0^T e\Phi(t)\, dt$—arises from the electron's interaction with the quantized A_μ, not from field strengths ($\mathbf{E}$, $\mathbf{B}$) or derived invariants alone. Gauge-invariant quantities in QED (e.g., $\langle F_{\mu\nu}F^{\mu\nu}\rangle$) describe observable outcomes but do not dynamically generate the phase during transit, mirroring QM's limitation (Section 3). The generalized AB effect's evidence of continuous accumulation, upheld in QED's path-integral formalism, excludes instantaneous gauge-invariant mechanisms, as the photon-mediated potential remains the operative agent. Even if QED introduces additional invariants (e.g., photon correlation functions), their contributions are too small to account for ϕ_{AB}'s magnitude, leaving $\mathbf{A}$'s role irreducible.

5.5 Conclusion

QED does not alter the conclusion that gauge-invariant explanations of the AB effect are inadequate; rather, it deepens it. The correspondence with QM ensures that the phase's continuous accumulation persists, while QED's quantized framework—despite proposals like Marletto and Vedral's or Vaidman's—reaffirms the necessity of potentials over field-derived invariants. The AB effect, in both QM and QED, reveals a reality where A_μ governs the electron's phase evolution, excluding gauge-invariant accounts that demand instantaneity at interference. This convergence invites a philosophical reassessment of potentials' ontological status, pursued in Section 6, as fundamental to quantum theory's explanatory structure.

6 Reality of potentials in the Lorenz gauge

The critique in the preceding sections has exposed the profound shortcomings of gauge-invariant explanations for the AB effect, highlighting their dependence on nonlocal and discontinuous mechanisms that contravene special relativity and the local essence of quantum interactions. This failure compels us to reconsider the role of electromagnetic potentials, long relegated to the status of mathematical conveniences due to their gauge dependence. In this section, I argue that the continuous accumulation of the AB phase—a central finding of this analysis—demands that electromagnetic potentials possess physical significance as direct representations of the underlying physical state, a role that fields and gauge-invariant quantities cannot fulfill. Furthermore, I argue that this significance is most coherently embodied in a single, preferred gauge—the Lorenz gauge—owing to its unique alignment with locality, Lorentz covariance, and causality (see Mulder, 2021 for a similar proposal).

6.1 The Physical Significance of Electromagnetic Potentials

The cornerstone of this paper's analysis is the revelation that the AB phase accumulates continuously as the electron traverses its path, a process that gauge-invariant explanations cannot adequately describe. This insight, drawn from Sections 3 through 5, fundamentally implies that electromagnetic potentials must be physically real, as fields like $\mathbf{E}$ and $\mathbf{B}$, and other gauge-invariant quantities such as $\rho = |\psi|^2$ and $\mathbf{v} = \frac{1}{m}(\nabla S - e\mathbf{A})$, fail to capture this dynamic evolution.

Consider the evidence: Section 3 established that gauge-invariant approaches—relying on the magnetic field $\mathbf{B}$ or phase holonomies—attribute the AB phase shift to an instantaneous emergence at the point of interference. This requires a nonlocal action from the solenoid's magnetic flux, despite the electron's confinement to a field-free region, contradicting special relativity's causal limits. Such explanations cannot account for a phase that builds progressively, as they tie the effect to the field's state at a single moment rather than the electron's journey.

Section 4's exploration of the generalized AB effect deepens this critique. When the magnetic flux varies with time, the phase shift is proportional to the flux's time average over the electron's travel duration, not merely its value at interference. This continuous accumulation is starkly evident in the proposed experiment: if the solenoid current is turned off before the beams overlap, a gauge-invariant account predicts no shift (since $\mathbf{B} = 0$ at that instant), yet quantum mechanics predicts a persistent effect tied to the potential's prior influence. The fields $\mathbf{E}$ and $\mathbf{B}$—zero or negligible outside the solenoid throughout much of the electron's path—offer no mechanism for this gradual phase build-up. Similarly, gauge-invariant quantities like ρ and $\mathbf{v}$ remain unchanged by the flux before interference, only shifting discontinuously upon overlap, as dictated by nonlocal conditions like $\oint \mathbf{v} \cdot d\mathbf{r} = 2\pi n - e\Phi$. This discontinuity clashes with the smooth, path-dependent phase evolution observed, underscoring their inadequacy.

By contrast, the electromagnetic potentials $A_\mu = (-A_0, \mathbf{A})$ naturally account for this continuous process. In the Schrödinger equation,

$$i\frac{\partial\psi}{\partial t} = -\frac{1}{2m}(\nabla - ie\mathbf{A})^2\psi + eA_0\psi, \tag{47}$$

the potentials couple directly to the wave function, inducing a phase $\Delta\phi = e\int_L A_\mu dx^\mu$ that accumulates along the electron's trajectory. In the magnetic AB effect, the vector potential $\mathbf{A}$ outside the solenoid, though yielding $\mathbf{B} = \nabla \times \mathbf{A} = 0$, drives a phase difference

$$\phi_{AB} = e\oint \mathbf{A} \cdot d\mathbf{r} \tag{48}$$

proportional to the enclosed flux, observable via interference. This phase emerges not from the fields, which vanish locally, but from the potential's persistent presence, integrated over the path. The generalized case (Section 2.2) reinforces this: a time-dependent flux $\Phi(t)$ yields

$$\phi_{AB} = \frac{1}{T}\int_0^T e\Phi(t)\,dt, \tag{49}$$

reflecting the potential's evolving influence over time, not an abrupt field effect. Only A_μ can mediate this continuous interaction, as it tracks the electron's state across its entire journey, unlike fields or gauge-invariant quantities confined to static or instantaneous descriptions.

This continuous phase accumulation thus necessitates that A_μ represents something physically real that affects the phase evolution of the electron, which determines the interference outcome. Fields like $\mathbf{B}$ describe forces, absent here, while ρ and $\mathbf{v}$ keep unchanged by the flux until overlap. The potential, however, actively shapes the phase of the wave function of the electron, even in field-free regions. Section 3's Madelung analysis confirms this limitation: without invoking $\mathbf{A}$, the nonlocal quantization condition fails to explain the phase's origin, leaving a gap that only the potential fills. Section 5's QED treatment further supports this, as gauge-invariant quantities (e.g., interaction energy) cannot eliminate A_μ's role for non-closed paths, affirming its necessity across frameworks.

If A_μ were merely a mathematical tool, its influence would be reducible to fields or derivable from gauge-invariant terms alone—yet the AB effect defies this reduction. The phase's continuous build-up in a field-free region, which depends on the potential's line integral, suggests A_μ encodes a physical reality beyond $\mathbf{E}$ and $\mathbf{B}$. In classical mechanics, potentials lack such status, as motion depends solely on forces; in quantum mechanics, however, the Schrödinger equation elevates A_μ to a fundamental driver of the electron's state, irreducible to field effects. The AB effect's ontology—where the electron's interference reflects a history of potential interactions—thus demands that electromagnetic potentials be physically real.

6.2 The Necessity of One True Gauge

Accepting that electromagnetic potentials are real confronts us with their gauge dependence: under $A_\mu \to A_\mu - \partial_\mu \Lambda$, paired with $\psi \to e^{-ie\Lambda}\psi$, observables remain invariant, suggesting arbitrariness. Yet, if A_μ is physically real, its representation must be determinate, implying the existence of one "true gauge" where it uniquely reflects this reality (Maudlin, 1998). The continuous phase accumulation reinforces this need: a process so intimately tied to A_μ's local values cannot be left ambiguous across multiple gauges without undermining its physical basis.

Gauge invariance ensures empirical equivalence, but ontological clarity requires a single, consistent form for A_μ. If every gauge were equally real, the potential at a point x—e.g., $\mathbf{A}(x)$ versus $\mathbf{A}(x) - \nabla\Lambda(x)$—would yield distinct descriptions of the same physical state, yet produce identical outcomes, creating an indeterminacy incompatible with a fundamental entity. In the AB effect, choosing $\mathbf{A} = 0$ in one path and a nonzero $\mathbf{A}$ in another is locally feasible, but globally reconciling these requires a transformation reflecting the flux—a task demanding a unified gauge. Multiple real gauges would either multiply realities excessively or necessitate nonlocal coordination, reintroducing the flaws of gauge-invariant accounts. A true gauge resolves this by fixing A_μ's form, ensuring its values correspond to the electron's continuous phase evolution without ambiguity.

A potential objection arises here: no observation could reveal the one true gauge (Maudlin, 1998). This critique suggests that the empirical underdetermination of A_μ undermines the claim of a single true gauge, as all observable quantities—e.g., the AB phase shift $\phi_{AB} = e \oint \mathbf{A} \cdot d\mathbf{r}$—remain invariant under gauge transformations, leaving no experimental means to distinguish one gauge from another. While this poses a legitimate challenge to direct verification, it does not invalidate the ontological necessity of a true gauge. The objection conflates empirical accessibility with physical reality: the fact that we cannot measure A_μ directly does not imply it lacks a determinate form. In quantum mechanics, an unknown quantum state is similarly unobservable in its full detail, yet we assign it ontological status based on its explanatory role. Likewise, A_μ's role in driving the continuous phase accumulation justifies its reality, even if its specific gauge-fixed values elude observation.

Moreover, the epistemological limitation is not unique to this proposal. Scientific theories often posit entities beyond direct empirical grasp—e.g., quarks—relying on their consistency with observable phenomena and theoretical coherence. The true gauge's identification rests on such grounds: the Lorenz gauge as the true gauge (see next subsection) is favored not by empirical discrimination but by its alignment with special relativity and locality, principles that constrain the state of reality. The AB effect's continuous phase, dependent on A_μ's path integral, demands a gauge that respects these constraints, not an arbitrary choice. While observation cannot pinpoint A_μ's form, the requirement of a consistent, causal description of reality—free of the nonlocality or discontinuity plaguing gauge-invariant accounts—necessitates a single gauge, undeterred by our inability to probe it directly. Thus, the epistemological objection, while highlighting a practical limit, does not negate the conceptual need for one true gauge to anchor the potentials' physical significance.

6.3 The Lorenz Gauge as the True Gauge

Having established the necessity of a single true gauge in which electromagnetic potentials A_μ represent the state of reality, we must now identify which gauge fulfills this role. I argue that the Lorenz gauge, defined by $\partial_\mu A^\mu = 0$, emerges as the *true gauge*—uniquely satisfying locality, Lorentz covariance, causality, and ontological completeness. This section dissects these virtues, cementing the Lorenz gauge as the determinate, relativistically consistent ontology for the AB effect.

6.3.1 Locality and Causal Propagation

The Lorenz gauge's primacy rests on its local, causal dynamics, inherently compatible with special relativity. Maxwell's equations under $\partial_\mu A^\mu = 0$ yield the wave equation,

$$\Box A^\mu = J^\mu, \tag{50}$$

where $\Box = \partial_\mu \partial^\mu$ is the Lorentz-invariant d'Alembertian, guarantees that changes in $A_\mu(\mathbf{r}, t)$ propagate at the speed of light, reflecting retarded influences from sources J^μ within the past light cone. This locality is pivotal: the generalized AB effect (Section 2.2) shows the phase accruing continuously along the electron's path, driven by A_μ's gradual action, not an abrupt effect of Φ at interference. Only in the Lorenz gauge does A_μ's evolution respect the causal structure of spacetime, ensuring that the phase accumulates via retarded interactions localized along the electron's trajectory. By contrast, the Coulomb gauge ($\nabla \cdot \mathbf{A} = 0$) defines $A_0 = -\int \frac{\rho(\mathbf{r}', t)}{|\mathbf{r}-\mathbf{r}'|} \, d^3\mathbf{r}'$, implying instantaneous dependence on distant charges, violating relativistic causality.

6.3.2 Lorentz Covariance

Lorentz covariance further elevates the Lorenz gauge as a physical necessity. As a four-vector, A_μ transforms consistently across inertial frames, with $\partial_\mu A^\mu = 0$ invariant, ensuring the AB phase's frame-independent predictions. For instance, shutting off the solenoid current before interference (Section 4.3) yields a persistent $\phi_{AB} = e\Phi$ due to retarded potentials—an effect the Lorenz gauge naturally encodes, while gauge-invariant accounts erroneously predict null shifts when $\mathbf{B} = 0$ at overlap. The Coulomb gauge, with its non-covariant A_0, disrupts this relativistic unity, failing to align the effect's spacetime dynamics with special relativity. Only the Lorenz gauge guarantees a consistent, universal description, a hallmark of a true physical ontology.

6.3.3 Determinacy and Uniqueness

The Lorenz gauge resolves gauge ambiguity, fixing A_μ uniquely. The general solution to Equation (50) is

$$A^\mu(x) = \int G(x - x') J^\mu(x') \, d^4x' + A^\mu_{\text{hom}}(x), \tag{51}$$

where $G(x-x')$ is the retarded Green's function and A^μ_{hom} satisfies $\Box A^\mu_{\text{hom}} = 0$. Imposing $\partial_\mu A^\mu = 0$ and boundary conditions (e.g., $A^\mu \to 0$ at infinity) constrains A^μ_{hom}: gauge shifts $\partial_\mu \Lambda$ must obey $\Box \Lambda = 0$, typically reducing to a constant, which is dynamically inert. In the AB effect, the solenoid's J^μ specifies a unique A_μ (e.g., $\mathbf{A} = \frac{\Phi(t-r/c)}{2\pi r} \hat{\boldsymbol{\theta}}$), determining the phase without residual freedom. This determinacy ensures A_μ as a unambiguous descriptor of reality, a prerequisite for its physical significance.

6.3.4 Ontological Sufficiency

Finally, the Lorenz gauge is ontologically sufficient; potentials in the Lorenz gauge can generate all gauge-invariant descriptions, rendering the inclusion of gauge-invariant quantities besides them in the ontology redundant. Given A_μ and the wave function ψ, all observables follow: $\rho = |\psi|^2$, $\mathbf{v} = \frac{1}{m}(\nabla S - e\mathbf{A})$, and $\mathbf{E}$, $\mathbf{B}$ via $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$, showing that they are mere manifestations of the underlying state defined by A_μ and ψ. The AB phase $\phi_{AB} = e\oint_C \mathbf{A} \cdot d\mathbf{r}$ emerges directly, rendering entities like Φ redundant. The Lorenz gauge, by fixing A_μ in a physically consistent form, thus provides a complete and economical description, eliminating the need for supplementary gauge-invariant constructs.

To sum up, the Lorenz gauge stands as the true gauge—not a convention, but the unique embodiment of A_μ's reality. Its locality, covariance, determinacy, and sufficiency align with the AB effect's demands, as Sections 4-5 affirm, offering a testable, coherent framework. This choice resolves the gauge problem decisively, paving the way for probing A_μ's deeper nature in Section 6.4, and solidifying its primacy in quantum interactions.

6.4 Nonzero Photon Mass and the Proca Equation

The argument thus far assumes a massless photon, consistent with standard electromagnetism and QED, where the Lorenz gauge fixes A_μ for a gauge-invariant theory with $\mathbf{E}$ and $\mathbf{B}$ as physical fields. However, an intriguing possibility arises if the photon possesses a nonzero mass: the Lorenz gauge then aligns with the Lorenz condition for the Proca equation, the simplest framework for a massive photon, potentially enhancing the physical reality of the so-called "gauge" potentials.

In standard QED, the photon is massless, and the action $S = \int(-\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + J^\mu A_\mu)\, d^4x$ permits gauge transformations $A_\mu \to A_\mu - \partial_\mu \Lambda$, leaving A_μ indeterminate. The Proca theory introduces a mass term with the action:

$$S = \int \left(-\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \frac{1}{2}m^2 A_\mu A^\mu + J^\mu A_\mu \right) d^4x \tag{52}$$

Varying this with respect to A_μ yields the field equation:

$$\partial_\mu F^{\mu\nu} + m^2 A^\nu = J^\nu \tag{53}$$

Taking the divergence, since $\partial_\mu \partial_\nu F^{\mu\nu} = 0$ due to antisymmetry, we obtain:

$$m^2 \partial_\nu A^\nu = 0 \tag{54}$$

For $m \neq 0$, this enforces:

$$\partial_\nu A^\nu = 0 \quad (55)$$

which is the Lorenz gauge condition, now a necessity rather than a choice. Then, the Proca equation becomes:

$$\Box A^\mu - m^2 A^\mu = J^\mu \quad (56)$$

a massive Klein-Gordon equation, where A_μ propagates as a physical field with three degrees of freedom (two transverse and one longitudinal), rather than the two polarizations of a massless photon. With boundary conditions (e.g., $A_\mu \to 0$ at infinity), this fixes the potentials A_μ uniquely. This determinacy aligns with the AB effect's demand for a local, causal A_μ (Section 6.2), eliminating the need to select a gauge manually.

For the AB effect, this shift has profound implications: the vector potential $\mathbf{A}$ outside the solenoid, driving the phase $\phi_{AB} = e \oint \mathbf{A} \cdot d\mathbf{r}$, becomes a tangible field, not a gauge artifact. The mass term ties A_μ directly to its sources (J^μ), eliminating the ambiguity of gauge choice, as $\partial_\mu A^\mu = 0$ is now a dynamic constraint rather than a freedom-reducing condition. The continuous phase accumulation observed in the generalized AB effect (Section 2) persists, with $\mathbf{A}$'s evolution governed by Equation (56), its range limited by m_γ^{-1} but still locally influencing the electron in the field-free region.

This nonzero mass enhances the potentials' reality in two key ways. First, the loss of gauge invariance elevates A_μ from a mathematical intermediary to a physical entity with measurable properties, akin to a massive particle's field. In the AB context, the phase shift remains observable, but $\mathbf{A}$'s contribution is no longer reducible to $\mathbf{B}$ alone; the longitudinal component, absent in the massless case, adds a physical degree of freedom directly tied to A_μ. Second, the Proca framework aligns with the Lorenz gauge's virtues—locality and retarded propagation—while removing the need to justify a "true gauge" against gauge freedom, as the mass inherently fixes A_μ. Experimental limits on photon mass are stringent ($m_\gamma < 10^{-18}\,\mathrm{eV/c^2}$). Yet, if a tiny mass exists—perhaps from beyond-standard-model physics—the potentials' enhanced reality in the Lorenz condition for the Proca equation strengthens their ontological primacy, offering a speculative yet coherent extension of this paper's argument.

6.5 Conclusion

To sum up, the continuous accumulation of the AB phase necessitates that electromagnetic potentials possess physical significance, a role fulfilled uniquely by the Lorenz gauge through its alignment with locality, Lorentz covariance, and causality. The potentials A_μ represent the electromagnetic state of the system, driving the electron's phase evolution in a manner irreducible to fields or gauge-invariant quantities, as evidenced across QM and QED (Sections 5 and 6). The Lorenz gauge's determinacy resolves gauge ambiguity, while a nonzero photon mass, via the Proca equation, could further concretize A_μ's reality by eliminating gauge freedom entirely. This analysis thus establishes electromagnetic potentials in the Lorenz gauge as the fundamental reality underlying the AB effect, challenging the primacy of gauge invariance and offering a coherent ontology for quantum interactions.

7 The Ontological Status of Potentials in the Lorenz Gauge

Having established A_μ in the Lorenz gauge embodies the physical reality that drives the AB effect's phase evolution, we confront a deeper question: what state of being do these potentials represent? If A_μ, fixed by $\partial_\mu A^\mu = 0$, is not a mere intermediary but the very essence of the electromagnetic influence, its ontological status demands scrutiny. Here, I propose that A_μ in the Lorenz gauge manifests the electromagnetic state of reality—a dynamic configuration shaping quantum interactions—and explore its potential identity as the classical precursor to a photon wave function in the position representation, explicitly tied to A_μ itself. This inquiry bridges the AB effect's empirical demands with a philosophical reimagining of quantum ontology, probing the nature of light's being.

7.1 Potentials as the Electromagnetic State

The AB effect illuminates A_μ's primacy: the phase $\phi_{AB} = e \oint_C \mathbf{A} \cdot d\mathbf{r}$ emerges not from local field strengths ($\mathbf{E}$, $\mathbf{B}$) but from the potentials' continuous presence along the electron's path, a process Sections 2–5 affirm as local and dynamic. In the Lorenz gauge, A_μ satisfies $\Box A^\mu = J^\mu$, propagating as a retarded field from sources J^μ, its four-vector form encoding the electromagnetic environment's influence. I posit that A_μ represents the *electromagnetic state*—a physical reality distinct from derived fields, capturing the potentiality of interaction rather than its realized forces. In the generalized AB effect (Section 2.2), the phase $\phi_{AB} = e \oint_C \mathbf{A}(\mathbf{r}, t) \cdot d\mathbf{r}$ reflects A_μ's temporal evolution, a history of influence irreducible to instantaneous Φ. Philosophically, this elevates A_μ beyond a mathematical scaffold: it is the state of being through which the quantum world communes with electromagnetism, fixed uniquely by the Lorenz gauge's causal clarity (Section 6.3).

7.2 The Photon Wave Function

Interpreting A_μ as the electromagnetic state, fixed in the Lorenz gauge, prompts an exploration of its quantum nature through a position-space photon wave function. Unlike non-relativistic quantum mechanics, where wave functions directly yield position probabilities, photons—due to their massless, relativistic, and non-localizable nature—require a generalized approach.

Here, I define $\psi_\mu(\mathbf{r},t) = \langle 0|\hat{A}_\mu(\mathbf{r},t)|\psi\rangle$ as the definitive representation of a single photon's state in space, prove its completeness and uniqueness, demonstrate its role in explaining quantum phenomena like the double-slit experiment, and establish its superiority over alternatives like the Riemann-Silberstein (RS) wave function, culminating in its foundational status within QED.

7.2.1 Definition of the Photon Wave Function

To rigorously define the photon wave function, we begin with the quantized electromagnetic field in quantum field theory (QFT), where the four-potential operator $\hat{A}_\mu(\mathbf{r},t)$ serves as the cornerstone of photon dynamics. In the Heisenberg picture, and adopting natural units ($c = \hbar = 1$), this operator is expressed as:

$$\begin{aligned}\hat{A}_\mu(\mathbf{r},t) &= \int \frac{d^3k}{(2\pi)^{3/2}} \sum_{\lambda=1}^{2} \frac{1}{\sqrt{2\omega_k}} [\epsilon_\mu(\mathbf{k},\lambda) a(\mathbf{k},\lambda) e^{i(\mathbf{k}\cdot\mathbf{r}-\omega_k t)} \\ &+ e^*_\mu(\mathbf{k},\lambda) a^\dagger(\mathbf{k},\lambda) e^{-i(\mathbf{k}\cdot\mathbf{r}-\omega_k t)}], \end{aligned} \tag{57}$$

where $\omega_k = |\mathbf{k}|$ ($\hbar = c = 1$), $\epsilon_\mu(\mathbf{k},\lambda)$ are transverse polarization vectors ($\lambda = 1, 2$) satisfying $k^\mu \epsilon_\mu = 0$, and $a(\mathbf{k},\lambda)$, $a^\dagger(\mathbf{k},\lambda)$ are annihilation and creation operators. Longitudinal and scalar modes ($\lambda = 0, 3$) are excluded via the Lorenz gauge condition.

The single-photon state is constructed in the Fock space as:

$$|\psi\rangle = \int d^3k \sum_{\lambda=1}^{2} f_\lambda(\mathbf{k}) \hat{a}^\dagger_\lambda(\mathbf{k})|0\rangle, \tag{58}$$

where $f_\lambda(\mathbf{k})$ is a complex amplitude function specifying the photon's momentum and polarization distribution, subject to the normalization condition:

$$\langle\psi|\psi\rangle = \int d^3k \sum_\lambda |f_\lambda(\mathbf{k})|^2 = 1, \tag{59}$$

ensuring a single-particle state. This $|\psi\rangle$ encapsulates the photon's quantum state in momentum space, with $f_\lambda(\mathbf{k})$ fully determining its properties. Now a position-space photon wave function can be defined as

$$\psi_\mu(\mathbf{r},t) = \langle 0|\hat{A}_\mu(\mathbf{r},t)|\psi\rangle = \int \frac{d^3k}{(2\pi)^{3/2}} \sum_{\lambda=1}^{2} \frac{f(\mathbf{k})}{\sqrt{2\omega_k}} \epsilon_\mu(\mathbf{k},\lambda) e^{i(\mathbf{k}\cdot\mathbf{r}-\omega_k t)}. \tag{60}$$

This $\psi_\mu(\mathbf{r},t)$ is a four-vector function encoding the photon's momentum and polarization distribution in position space, akin to a classical potential. It satisfies $\Box A_\mu = 0$ and $\partial^\mu A_\mu = 0$, mirroring a free photon's dynamics.

Unlike non-relativistic wave functions, $|\psi_\mu(\mathbf{r},t)|^2$ isn't a probability density—photons lack a position operator due to their massless nature—but $\psi_\mu(\mathbf{r},t)$ generates fields for QED observables, making it a generalized wave function tailored to the photon's relativistic and gauge-dependent essence.

7.2.2 Proof of Completeness and Uniqueness

To establish $\psi_\mu(\mathbf{r},t)$ as the definitive photon wave function, we must demonstrate its completeness—its capacity to encapsulate all physical information of the single-photon state $|\psi\rangle$ in position space—and its uniqueness—its status as the singularly appropriate representation within QED's framework. Let us proceed with rigor, building from first principles and addressing potential challenges.

Completeness begins with recognizing that $|\psi\rangle = \int d^3k \sum_\lambda f_\lambda(\mathbf{k})\hat{a}^\dagger_\lambda(\mathbf{k})|0\rangle$ fully specifies the photon's quantum state in momentum space, with $f_\lambda(\mathbf{k})$ encoding its momentum distribution and polarization content across two transverse modes ($\lambda = 1, 2$), normalized via $\int d^3k \sum_\lambda |f_\lambda(\mathbf{k})|^2 = 1$. Second, the transition to position space via $\psi_\mu(\mathbf{r},t) = \langle 0|\hat{A}_\mu(\mathbf{r},t)|\psi\rangle$ preserves this entirety. $\psi_\mu(\mathbf{r},t)$ retains all of $f_\lambda(\mathbf{k})$ and $\epsilon^{(\lambda)}_\mu(k)$, mapped into spacetime and suggesting no information is lost.

To verify, consider the physical content $\psi_\mu(\mathbf{r},t)$ delivers. The electromagnetic fields follow: $\hat{\mathbf{E}}(\mathbf{r},t) = -\nabla\hat{A}_0(\mathbf{r},t) - \partial_t\hat{\mathbf{A}}(\mathbf{r},t)$, $\hat{\mathbf{B}}(\mathbf{r},t) = \nabla\times\hat{\mathbf{A}}(\mathbf{r},t)$. For $\psi_\mu(\mathbf{r},t) = (\psi, \psi_0)$, in Lorenz gauge ($\partial^\mu\hat{A}_\mu = 0$), typically $\psi_0 = 0$ for transverse photons, so:

$$\langle 0|\hat{\mathbf{E}}(\mathbf{r},t)|\psi\rangle = -\partial_t\psi(\mathbf{r},t), \quad \langle 0|\hat{\mathbf{B}}(\mathbf{r},t)|\psi\rangle = \nabla\times\psi(\mathbf{r},t). \tag{61}$$

For potential effects like the AB effect, $\psi_\mu(\mathbf{r},t)$ directly provides $A_\mu(\mathbf{r},t)$'s quantum analog, with ϕ_{AB} derived from its classical counterpart. Could additional information—say, higher-order correlations—be missing? Multipoint correlators (e.g., $\langle\psi|\hat{A}_\mu(\mathbf{r},t)\hat{A}_\nu(\mathbf{r}',t)|0\rangle$) involve multi-photon states, irrelevant for a single photon, where $|\psi\rangle$'s Fock-space definition ensures all properties are encoded in $f_\lambda(\mathbf{k})$, fully transferred to $\psi_\mu(\mathbf{r},t)$. Thus, completeness holds: $\psi_\mu(\mathbf{r},t)$ exhausts the single-photon state's spatial description.

Uniqueness demands that $\psi_\mu(\mathbf{r},t)$ be the sole appropriate position-space representation. Its basis in $\hat{A}_\mu(\mathbf{r},t)$—QED's fundamental field operator—is the linchpin. Alternative operators might be proposed—e.g., the field-strength tensor $\hat{F}_{\mu\nu}(\mathbf{r},t) = \partial_\mu\hat{A}_\nu(\mathbf{r},t) - \partial_\nu\hat{A}_\mu(\mathbf{r},t)$. Compute:

$$\langle 0|\hat{F}_{\mu\nu}(\mathbf{r},t)|\psi\rangle = \partial_\mu\psi_\nu(\mathbf{r},t) - \partial_\nu\psi_\mu(\mathbf{r},t), \tag{62}$$

a derivative of $\psi_\mu(\mathbf{r},t)$, not an independent wave function. Any $\hat{F}_{\mu\nu}$-based formulation (e.g., $\mathbf{F} = \mathbf{E} + i\mathbf{B}$) is thus subordinate, lacking $A_\mu(\mathbf{r},t)$'s primacy in potential-driven effects. Could another operator, say a scalar field, serve? Photons have vector nature (spin-1), and QED's structure ties them to $\hat{A}_\mu(\mathbf{r}, t)$, with no scalar or tensor alternative matching gauge and polarization constraints. Gauge transformations ($\hat{A}_\mu \to \hat{A}_\mu + \partial_\mu \chi$, $\psi_\mu \to \psi_\mu + \partial_\mu \chi$) alter $\psi_\mu(\mathbf{r}, t)$'s form, but QED's gauge-invariant observables (e.g., ϕ_{AB}) remain unchanged, and Lorenz gauge fixes a consistent representation without loss of generality—other gauges (e.g., Coulomb) are equivalent via transformation, not distinct.

Objections might arise: does $\psi_\mu(\mathbf{r}, t)$'s non-localizability (no $|\psi_\mu(\mathbf{r}, t)|^2$ as probability) undermine completeness? For massive particles, position operators exist, but photons' massless nature and helicity preclude this—Newton-Wigner fails, producing unphysical states. $\psi_\mu(\mathbf{r}, t)$ adapts by defining fields and potentials, sufficient for QED's statistical predictions, aligning with the photon's relativistic essence rather than a flaw. Could multiple wave functions coexist? Any alternative must replicate $|\psi\rangle$'s information and QED's observables; since $\psi_\mu(\mathbf{r}, t)$ does so via $\hat{A}_\mu(\mathbf{r}, t)$, additional forms (e.g., ad hoc scalars) are redundant or inconsistent. Thus, uniqueness holds: $\psi_\mu(\mathbf{r}, t)$ is the canonical, singular bridge from momentum to position space in QED.

7.2.3 Explaining the Double-Slit Experiment as an Example

To concretely illustrate the physical significance of the photon wave function $\psi_\mu(\mathbf{r}, t) = \langle 0|\hat{A}_\mu(\mathbf{r}, t)|\psi\rangle$, we analyze the canonical double-slit experiment. This example demonstrates how the quantum electromagnetic potential A_μ, encoded in $\psi_\mu(\mathbf{r}, t)$, governs interference phenomena.

Consider a single photon incident on a double-slit apparatus. The quantum state $|\psi\rangle$ is a superposition of two paths:

$$|\psi\rangle = \frac{1}{\sqrt{2}} \left(|\psi_1\rangle + |\psi_2\rangle\right), \tag{63}$$

where $|\psi_i\rangle$ corresponds to the photon passing through slit i. The photon wave function becomes:

$$\psi_\mu(\mathbf{r}, t) = \frac{1}{\sqrt{2}} \left(\psi_\mu^{(1)}(\mathbf{r}, t) + \psi_\mu^{(2)}(\mathbf{r}, t)\right), \tag{64}$$

with $\psi_\mu^{(i)}(\mathbf{r}, t) = \langle 0|\hat{A}_\mu(\mathbf{r}, t)|\psi_i\rangle$ describing the potential field configuration for path i.

Assuming slits at positions $\mathbf{r}_1$ and $\mathbf{r}_2$, the wave functions propagate as spherical waves:

$$\psi_\mu^{(i)}(\mathbf{r}, t) \propto \frac{e^{i(k|\mathbf{r}-\mathbf{r}_i|-\omega t)}}{|\mathbf{r}-\mathbf{r}_i|} \epsilon_\mu(\mathbf{k}), \tag{65}$$

where $\epsilon_\mu(\mathbf{k})$ is the polarization vector. The total wave function at the screen is:

$$\psi_\mu(\mathbf{r}, t) \propto \frac{e^{ikL_1}}{L_1} \epsilon_\mu + \frac{e^{ikL_2}}{L_2} \epsilon_\mu, \tag{66}$$

where $L_i = |\mathbf{r} - \mathbf{r}_i|$. For $L_1 \approx L_2 \equiv L$, the path difference $\Delta L = L_1 - L_2$ introduces a phase difference:

$$\psi_\mu \propto \frac{2e^{ikL}}{L}\epsilon_\mu \cos\left(\frac{k\Delta L}{2}\right). \tag{67}$$

The detection probability can be calculated using the following gauge-invariant prescription (Glauber, 1963; Mandel and Wolf, 1995):[2]

$$P(\mathbf{r}, t) \propto \langle\psi|\hat{\mathbf{E}}^-(\mathbf{r}, t) \cdot \hat{\mathbf{E}}^+(\mathbf{r}, t)|\psi\rangle, \tag{68}$$

where $\hat{\mathbf{E}}^\pm = -\partial_t \hat{\mathbf{A}}^\pm - \nabla \hat{A}_0^\pm$ are the positive/negative frequency components of the electric field operator. Substituting $\hat{\mathbf{A}}^\pm$ with ψ_μ, we compute:

$$\hat{\mathbf{E}}^- \cdot \hat{\mathbf{E}}^+ \propto |\partial_t \psi_\mu|^2 + |\nabla \psi_\mu|^2. \tag{69}$$

For the superposed wave function, this yields:

$$P(\mathbf{r}) \propto \cos^2\left(\frac{k\Delta L}{2}\right), \tag{70}$$

reproducing the observed interference fringes. Critically, this result emerges from the superposition of potentials (ψ_μ), not the fields $\mathbf{E}$ or $\mathbf{B}$.

To sum up, the double-slit experiment demonstrates that interference patterns arise from the superposition of electromagnetic potentials A_μ (ψ_μ). This reaffirms the thesis that potentials in the Lorenz gauge are ontologically fundamental, while gauge invariance reflects descriptive redundancy, not physical equivalence.

7.2.4 Comparison with Alternatives

To grasp the full measure of $\psi_\mu(\mathbf{r}, t)$'s significance, we must turn our gaze to the landscape of alternatives that have sought to represent the photon's state in space, each offering a glimpse of insight yet falling short of the comprehensive vision embodied in our formulation.

[2] A rigorous analysis may show that no standard probability density $\rho(\mathbf{r}, t)$ and current $\mathbf{j}(\mathbf{r}, t)$—satisfying positivity, normalization to unity ($\int \rho d^3x = 1$), continuity ($\partial_t \rho + \nabla \cdot \mathbf{j} = 0$), and Lorentz covariance—can exist for a single photon (Newton and Wigner, 1949; Bialynicki-Birula, 1996; Sebens, 2019). Suppose such a $\rho = \langle\psi|\hat{\rho}|\psi\rangle$ exists, with $\int \hat{\rho} d^3x = \hat{N}$ (number operator). Covariance suggests $j^\mu = (c\rho, \mathbf{j}) \propto k^\mu F(k \cdot x)$, but integrating yields energy $P^0 = E/c$, not 1, and locality of $\hat{\rho}$ conflicts with the non-local $\hat{N}$ due to light-cone correlations of massless fields (e.g., $\langle A^\mu(x) A^\nu(y)\rangle \sim 1/(x-y)^2$). Thus, photons' massless, spin-1 nature and gauge invariance preclude a local, normalizable probability density and conserved current, making standard definitions impossible. However, a Lorentz-covariant probability rule is still needed for predictions for the experimental setups where the speeds of the detectors in the laboratory frame are relativistic. This is an important challenge for theorists.

Consider first the momentum-space state $|\psi\rangle$, articulated as an integral over creation operators weighted by $f_\lambda(\mathbf{k})$ and rooted in the pristine clarity of quantum field theory's Fock space. Its completeness in momentum space is undeniable, offering a direct conduit to QED's computational machinery, whether for scattering processes or multi-photon states, and its normalization flows naturally from the photon's particle nature. Yet this purity comes at a cost: it remains silent on the photon's spatial presence, unable to manifest $A_\mu(\mathbf{r}, t)$, $\mathbf{E}(\mathbf{r}, t)$, or $\mathbf{B}(\mathbf{r}, t)$ without the intervention of an operator like $\hat{A}_\mu(\mathbf{r}, t)$. For phenomena tethered to position—interference patterns or potential-driven phase shifts—this silence is a void that $\psi_\mu(\mathbf{r}, t)$ fills, serving as the necessary bridge from the abstract expanse of momentum to the tangible theater of space, rendering it not a rival but a complementary necessity, elevated by its spatial articulation.

Next, consider the Riemann-Silberstein wave function, a construct that elegantly melds the electric and magnetic fields into a single complex entity, expressed as $\langle\psi|\hat{\mathbf{E}}(\mathbf{r}, t) + i\hat{\mathbf{B}}(\mathbf{r}, t)|0\rangle$, which in momentum space unfolds as an integral over wave vectors, weighted by polarization and field amplitudes, and in our terms emerges directly from $\psi_\mu(\mathbf{r}, t)$ as $-\partial_t\psi(\mathbf{r}, t) + i\nabla \times \psi(\mathbf{r}, t)$ when the scalar potential vanishes. Its allure lies in its gauge invariance—a quality that frees it from the shackles of arbitrary gauge choices—and its intuitive resonance with classical optics, where its magnitude squared approximates energy density. Yet this elegance is marred by a profound limitation: it lacks Lorentz covariance, for under a relativistic boost, the electric and magnetic components transform in a manner that disrupts its form (e.g., the z-component shifts as $\gamma(E_z - vB_y) + i\gamma(B_z + vE_y)$), rendering it inadequate for a fully relativistic theory. More critically, it falters in regions where fields vanish but potentials persist, as in the AB effect, where $\mathbf{F}$ collapses to zero despite the photon's interaction with A_μ, a domain where $\psi_\mu(\mathbf{r}, t)$ thrives by embracing the four-potential's primacy. Confined to massless photons, it lacks the adaptability of $\psi_\mu(\mathbf{r}, t)$, which can extend to hypothetical massive scenarios, positioning our wave function as the more foundational descriptor.

Turning to the Landau-Peierls wave function, we encounter a historical attempt to cast the photon in a semi-classical mold, defined as a three-vector integral over momentum with a modified normalization factor, $1/\sqrt{2\omega_k}$, aiming to align its magnitude squared with a probability density akin to massive particles. Its intent—to bridge quantum and classical intuitions—carries a certain nostalgic charm, reflecting early efforts to grapple with the photon's elusive nature. Yet this ambition stumbles: the normalization disrupts Lorentz invariance, the lack of grounding in QFT's operator formalism leaves it adrift from modern rigor, and its pursuit of localization clashes with the photon's intrinsic delocalization, yielding inconsistent probabilities. Against this, $\psi_\mu(\mathbf{r}, t)$ stands as a beacon of consistency, rooted in $\hat{A}_\mu(\mathbf{r}, t)$, preserving relativistic symmetry, and eschewing false promises of position density, instead channeling its spatial insights through QED's robust framework, subsuming the Landau-Peierls field aspects with greater fidelity.

The Newton-Wigner approach offers yet another lens, one forged for massive particles, seeking to define position eigenstates whose wave functions promise a direct probability interpretation. For electrons or protons, this vision succeeds, anchoring their states in a spatial certainty that resonates with intuition. But for photons, massless and bound to light's velocity, this dream unravels: the absence of a rest frame and the fixed helicity of ± 1 defy localization, with boosts mixing helicities and introducing spurious longitudinal modes, rendering the approach untenable. $\psi_\mu(\mathbf{r}, t)$ sidesteps this quagmire, embracing the photon's relativistic essence as a delocalized entity, its four-vector form aligning with QFT's truths rather than forcing an ill-fitting classical mold, thus emerging as the truer representation for light's quantum nature.

Finally, we contemplate alternative gauge choices, such as the Coulomb gauge, where $\hat{A}_0 = 0$ and $\nabla \cdot \hat{\mathbf{A}} = 0$, reducing $\psi_\mu(\mathbf{r}, t)$ to a spatial vector transverse by design. This choice simplifies certain analyses, stripping away the scalar potential and spotlighting the photon's transverse degrees of freedom with a clarity that can aid computation. Yet this simplicity is no departure from $\psi_\mu(\mathbf{r}, t)$'s essence; it is merely a transformation—a rephrasing of the same underlying state through a gauge shift like $\hat{A}'_\mu = \hat{A}_\mu + \partial_\mu \chi$—lacking the generality of the Lorenz gauge's four-vector scope, which retains flexibility for broader contexts. $\psi_\mu(\mathbf{r}, t)$ in its Lorenz form thus holds the higher ground, its covariance and adaptability eclipsing the narrower focus of Coulomb's spatial lens, with no loss of physical content.

In this reflective traversal, $\psi_\mu(\mathbf{r}, t)$ emerges not merely as a contender but as the philosophical and physical linchpin, its direct tie to $A_\mu(\mathbf{r}, t)$ unlocking potential-driven effects, its QFT foundation via $\hat{A}_\mu(\mathbf{r}, t)$ ensuring rigor, its Lorentz covariance aligning with relativity's demands, its adaptability extending beyond massless constraints, and its capacity to engender all field-based alternatives cementing its primacy. Where others falter—be it in covariance, scope, or fidelity to the photon's nature—$\psi_\mu(\mathbf{r}, t)$ stands resolute, the definitive spatial voice of the photon's quantum being.

7.2.5 Conclusion

$\psi_\mu(\mathbf{r}, t) = \langle 0|\hat{A}_\mu(\mathbf{r}, t)|\psi\rangle$ is arguably the definitive position-space photon wave function, uniquely suited for the AB effect's ontology and all single-photon phenomena. Its completeness—capturing $|\psi\rangle$'s full state in x-space—and uniqueness—rooted in $\hat{A}_\mu(\mathbf{r}, t)$ with no independent rivals—establish it as the fundamental representation. It underpins QED's explanation of phenomena like the double-slit experiment, AB effect, and polarization via its state and field amplitudes, surpassing alternatives despite its own gauge dependence and lack of direct probability density, reflecting the photon's nature. Within QED, $\psi_\mu(\mathbf{r}, t)$ is the real and complete spatial descriptor of a single photon, with no need for other formulations.

7.3 Quantum Origin of Classical Potentials

The classical electromagnetic four-potential $A_\mu = (-A_0, \mathbf{A})$ governs macroscopic fields via $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$, yet its quantum origin in QED reveals a deeper reality tied to the photon wave function ψ_μ and the structure of many-photon states. This subsection elucidates how A_μ emerges from the quantized field $\hat{A}_\mu$, bridging quantum and classical domains through coherent states while contrasting their behavior with entangled states. This analysis reinforces the critique of gauge-invariant paradigms (Sections 3–4) by demonstrating that A_μ's classical form is a statistical outcome of quantum dynamics, not a static invariant, aligning with the AB effect's demand for potential-driven phase accrual (Section 2).

Classically, A_μ is real, combining positive and negative frequencies. For a plane wave ($A_0 = 0$):

$$\mathbf{A}(\mathbf{r}, t) = \mathbf{A}_0 e^{i(\mathbf{k}\cdot\mathbf{r}-\omega t)} + \mathbf{A}_0^* e^{-i(\mathbf{k}\cdot\mathbf{x}-\omega t)}, \tag{71}$$

ensuring:

$$\mathbf{A}(\mathbf{r}, t) = \mathbf{A}^*(\mathbf{r}, t), \tag{72}$$

with fields $\mathbf{E} = -\partial_t \mathbf{A}$ and $\mathbf{B} = \nabla \times \mathbf{A}$ scaling with intensity $|\mathbf{A}_0|^2$, indicative of many photons.

The classical A_μ emerges in QED via many-photon coherent states, which mimic macroscopic fields. For a coherent state $|\alpha\rangle$, where $\hat{a}_\lambda(\mathbf{k})|\alpha\rangle = \alpha_\lambda(\mathbf{k})|\alpha\rangle$ and $\alpha_\lambda(\mathbf{k}) = \sqrt{N} f_\lambda(\mathbf{k})$ with mean photon number N:

$$\begin{aligned} \langle\alpha|\hat{A}_\mu(\mathbf{r}, t)|\alpha\rangle &= \sum_\lambda \int \frac{d^3k}{(2\pi)^3} \frac{1}{\sqrt{2\omega_k}} [\alpha_\lambda(\mathbf{k})\epsilon_\mu(\mathbf{k}, \lambda) e^{i(\mathbf{k}\cdot\mathbf{x}-\omega_k t)} \\ &+ \alpha_\lambda^*(\mathbf{k})\epsilon_\mu^*(\mathbf{k}, \lambda) e^{-i(\mathbf{k}\cdot\mathbf{x}-\omega_k t)}], \end{aligned} \tag{73}$$

simplifying to:

$$\langle\alpha|\hat{A}_\mu(\mathbf{r}, t)|\alpha\rangle = \sqrt{N}[\psi_\mu(\mathbf{r}, t) + \psi_\mu^*(\mathbf{r}, t)]. \tag{74}$$

As $N \to \infty$, this matches A_μ: $\sqrt{N}$ scales amplitude to classical strength, and ψ_μ^* adds negative frequencies, ensuring reality. Coherent states' phase coherence—unlike Fock states' zero expectation—bridges quantum ψ_μ to classical A_μ, reinforcing A_μ's origin as a statistical average over many photons, not a gauge-invariant flux.

Entangled many-photon states, however, resist this classical limit. For a two-photon entangled state, e.g.:

$$|\Psi\rangle = \frac{1}{\sqrt{2}} \left[\hat{a}^\dagger_{\lambda_1}(\mathbf{k}_1)\hat{a}^\dagger_{\lambda_2}(\mathbf{k}_2) + \hat{a}^\dagger_{\lambda_3}(\mathbf{k}_3)\hat{a}^\dagger_{\lambda_4}(\mathbf{k}_4) \right] |0\rangle, \tag{75}$$

the expectation is:

$$\langle\Psi|\hat{A}_\mu(\mathbf{r}, t)|\Psi\rangle = 0, \tag{76}$$

due to mode mismatch: $\hat{A}_\mu^{(+)}$ and $\hat{A}_\mu^{(-)}$ connect to states with differing photon numbers or momenta, lacking the phase alignment of coherent states. Entangled states, rich in quantum correlations (e.g., nonlocality), do not produce a classical A_μ, highlighting that classicality requires coherence, not merely photon number.

Thus, A_μ's quantum origin lies in $\hat{A}_\mu$'s expectation over coherent many-photon states built from ψ_μ, scaled by $\sqrt{N}$ to macroscopic reality. This process—dynamic and potential-driven—parallels the AB effect's phase accumulation via A_μ (Section 2), not static invariants like $F_{\mu\nu}$ or Φ, supporting QED's reinforcement of a potential-centric ontology over gauge-invariant alternatives critiqued in Section 4.

8 Implications for Gauge Theories

The foregoing analysis has elevated electromagnetic potentials A_μ in the Lorenz gauge from mere auxiliaries to the fundamental bearers of reality, their continuous mediation of the AB phase (Sections 2-5) and ontological primacy (Section 6) dismantling the sufficiency of gauge-invariant constructs like Φ or $F_{\mu\nu}$. Fixed by $\partial_\mu A^\mu = 0$, A_μ emerges as the dynamic essence of electromagnetic interactions—a potential-centric ontology that transcends the classical bias toward field strengths and redefines the photon's being as $\psi_\mu(\mathbf{r}, t) \propto A_\mu(\mathbf{r}, t)$ (Section 6.4). This shift reverberates beyond the AB effect's confines, casting a critical light on gauge theories writ large. In this section, we probe these ripples: first, reinterpreting the Berry phase's genesis through potential-like mediators (Section 8.1); then, questioning gauge-invariant accounts in non-Abelian frameworks like Yang-Mills (Section 8.2); next, critiquing static formulations of the Higgs mechanism (Section 8.3); and finally, proposing a gravitational AB effect to affirm spacetime's substantival reality in general relativity (Section 8.4). Together, these inquiries suggest that potentials, not their invariants, may unify the causal and ontological fabric of modern physics.

8.1 The Origin of the Berry Phase

The previous analysis establishes the continuous accumulation of the AB phase via electromagnetic potentials A_μ in the Lorenz gauge (Sections 2-6), bear intriguing implications for understanding the Berry phase—a geometric phase acquired by a quantum system under adiabatic evolution along a closed path in parameter space (Berry, 1984). Like the AB phase $\phi_{AB} = e \oint_C \mathbf{A} \cdot d\mathbf{r}$, the Berry phase $\gamma = i \oint_C \langle \psi(\mathbf{R}) | \nabla_{\mathbf{R}} \psi(\mathbf{R}) \rangle \cdot d\mathbf{R}$ is gauge-invariant, emerging from a cyclic process and observable in interference or topological phenomena. However, the critique of gauge-invariant explanations (Sections 3-4)—their nonlocality, discontinuity, and inability to trace the AB phase's dynamic genesis—prompts a parallel inquiry into the Berry phase's accumulation. Conventionally treated as a geometric effect tied to the curvature of an abstract connection $\mathbf{A}(\mathbf{R}) = i\langle \psi(\mathbf{R}) | \nabla_{\mathbf{R}} \psi(\mathbf{R}) \rangle$, the Berry phase's process of generation may similarly depend on a gauge-dependent entity whose local influence, akin to A_μ, drives its continuous buildup.

This perspective challenges the standard view that the Berry phase's gauge invariance fully encapsulates its physical reality. Just as the AB effect's phase requires the vector potential $\mathbf{A}$ over the static flux Φ (Section 6.1), the Berry phase's reliance on $\mathbf{A}(\mathbf{R})$ suggests that its accumulation might reflect a physical state—potentially a potential-like mediator—rather than a purely emergent geometric property. In systems where the Berry phase arises from electromagnetic interactions (e.g., via $\mathbf{A}$ in a magnetic field), this analogy is direct, and the advocacy for the Lorenz gauge (Section 6.3) could imply a preferred gauge for $\mathbf{A}(\mathbf{R})$ in relativistic contexts, such as QED. The rejection of nonlocal mechanisms (Section 3.3) further questions interpretations that attribute the Berry phase solely to global parameter-space topology, potentially overlooking the local dynamics of Hamiltonian evolution mediated by such potentials.

Future directions could explore this connection empirically and theoretically. Experiments analogous to those in Section 4.3—tracking phase accrual in time-varying parameter spaces (e.g., dynamic magnetic fields or synthetic gauge fields in condensed matter)—might test whether the Berry phase accumulates continuously via a gauge-dependent entity, mirroring the generalized AB effect (Section 2.2). Theoretically, extending the potential-centric ontology to non-Abelian Berry phases in Yang-Mills theories (Section 8.2) could unify these phenomena under a common framework, where gauge-fixed potentials underpin observable shifts. Philosophically, this aligns with the critique of gauge-invariant ontologies (Section 4), suggesting that the Berry phase's invariance might obscure a deeper reality tied to physical mediators, enriching its interpretation beyond geometry and resonating with the AB effect's transformative implications for quantum theory.

8.2 Non-Abelian Gauge Theories

The critique of gauge-invariant explanations for the AB effect, as unfolded in Sections 3 through 5, and the subsequent affirmation of electromagnetic potentials A_μ in the Lorenz gauge as the fundamental mediators of quantum interactions (Section 6), extend beyond the Abelian $U(1)$ framework of electromagnetism to resonate with non-Abelian gauge theories, such as those of the Yang-Mills type underlying the Standard Model (Berghofer et al, 2023). In the AB effect, the continuous accumulation of the phase shift, driven by A_μ rather than static invariants like Φ, necessitated a rejection of gauge-invariant accounts marred by nonlocality, discontinuity, and incompleteness. Non-Abelian gauge theories, characterized by richer symmetry groups (e.g., $SU(2)$ for weak interactions, $SU(3)$ for strong interactions), similarly involve gauge potentials $\mathcal{A}_\mu^a$ (where a indexes the group generators) that transform under local gauge symmetries, raising parallel questions about their physical reality versus derived field strengths $\mathcal{F}_{\mu\nu}^a$. This section examines whether the lessons from the AB effect—favoring a potential-centric ontology—apply to these theories, exploring potential analogs of the AB effect and their implications for gauge invariance.

8.2.1 Non-Abelian Gauge Structure and Potentials

In non-Abelian gauge theories, the vector potential $\mathcal{A}_\mu = \mathcal{A}_\mu^a T^a$ (with T^a the Lie algebra generators) transforms under local gauge transformations $g(x) \in G$ as $\mathcal{A}_\mu \to g\mathcal{A}_\mu g^{-1} - \frac{i}{e} g\partial_\mu g^{-1}$, where G is a non-Abelian group like $SU(N)$. The field strength $\mathcal{F}_{\mu\nu} = \partial_\mu \mathcal{A}_\nu - \partial_\nu \mathcal{A}_\mu - ie[\mathcal{A}_\mu, \mathcal{A}_\nu]$ is gauge-covariant but not invariant, reflecting the non-commutative nature of the group, unlike the Abelian $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$. Gauge-invariant quantities, such as $\mathrm{Tr}(\mathcal{F}_{\mu\nu}\mathcal{F}^{\mu\nu})$, are often prioritized in conventional accounts, paralleling the AB effect's focus on Φ. However, the AB critique (Section 3.3) suggests that such invariants may fail to capture the dynamic generation of physical effects if they overlook the potentials' role. In non-Abelian theories, phenomena like confinement in quantum chromodynamics (QCD) or electroweak symmetry breaking hint at potentials' influence beyond field strengths, prompting inquiry into whether a non-Abelian AB-like effect could reveal $\mathcal{A}_\mu^a$'s primacy, akin to A_μ in the Abelian case.

8.2.2 Non-Abelian Analogs of the AB Effect

Could a non-Abelian AB effect, where a charged particle's phase depends on $\mathcal{A}_\mu^a$ in a field-free region, parallel the Abelian case? In $SU(N)$ theories, a Wilson loop $\mathcal{W}_C = \mathrm{Tr}\left[\mathcal{P}\exp\left(ie\oint_C \mathcal{A}_\mu dx^\mu\right)\right]$ (where $\mathcal{P}$ denotes path ordering) is gauge-invariant and generalizes the Abelian phase factor $e^{ie\oint_C \mathbf{A}\cdot d\mathbf{r}}$. Theoretical proposals suggest such an effect: a particle in a representation of $SU(N)$ traversing a region with vanishing $\mathcal{F}_{\mu\nu}^a$ but nonzero $\mathcal{A}_\mu^a$ (e.g., around a chromomagnetic flux tube) could acquire a phase proportional to the loop integral, observable via interference. Unlike the Abelian AB effect, the non-commutative structure complicates the phase's locality, as $\mathcal{A}_\mu^a$'s components interact nonlinearly. Yet, Section 2.2's generalized AB effect, with its time-varying flux, suggests a dynamic phase accumulation could apply, raising the question: do gauge-invariant accounts (e.g., relying solely on $\mathcal{W}_C$) suffer from nonlocality akin to the Abelian case, where Φ's distant influence lacks a local mediator (Section 3.3)?

8.2.3 Nonlocality and Discontinuity Concerns

A gauge-invariant explanation of a non-Abelian AB effect, focusing on $\mathcal{W}_C$, might mirror the AB critique's nonlocality objection. The Wilson loop integrates $\mathcal{A}_\mu^a$ globally along C, implying the phase depends instantaneously on the flux configuration, potentially distant from the particle's path, without a clear local mechanism—akin to the solenoid's nonlocal action in Section 3.3. In a dynamic scenario, akin to $\Phi(t)$ in Section 2.2, the phase's evolution could depend on $\mathcal{A}_\mu^a(t)$'s spacetime profile, exacerbating this nonlocality unless retarded propagation constrains it, a complexity absent in static Abelian cases. Similarly, discontinuity arises: if the phase emerges only at interference, as in gauge-invariant AB accounts (Section 3.3), the lack of a continuous generation process—unlike the potential-driven $\phi_{AB} = \frac{1}{T}\int_0^T e\Phi(t)dt$—clashes with quantum mechanics' smooth dynamics, suggesting non-Abelian gauge-invariant explanations may inherit the same temporal abruptness critiqued earlier.

8.2.4 Incompleteness and Potential-Centric Alternatives

The incompleteness objection (Section 3.3) also applies: a gauge-invariant focus on $\mathcal{W}_C$ or $\mathcal{F}_{\mu\nu}^a$ may fail to elucidate how the phase accumulates, omitting $\mathcal{A}_\mu^a$'s dynamic role, much as Φ overlooks $\mathbf{A}$'s contribution in the AB effect. In non-Abelian theories, phenomena like gluon confinement or electroweak mass generation involve $\mathcal{A}_\mu^a$'s interactions, suggesting potentials drive physical effects beyond invariants. A potential-centric ontology, as in Section 6, could posit $\mathcal{A}_\mu^a$ in a fixed gauge as the fundamental entity, with a non-Abelian AB effect's phase reflecting continuous accumulation via $\oint_C \mathcal{A}_\mu^a dx^\mu$.

8.2.5 Broader Gauge Theory Implications

The AB effect's elevation of potentials A_μ over gauge-invariant quantities like Φ (Sections 6.1–6.4) reverberates through the edifice of non-Abelian gauge theories, unsettling the foundational reliance on invariants such as Wilson loops ($\mathcal{W}_C$) or field strengths ($\mathcal{F}_{\mu\nu}^a$). Just as the continuous phase accumulation in the generalized AB effect (Section 2.2) exposed the nonlocality and discontinuity of invariant-centric accounts (Sections 3–4), a potential non-Abelian AB analog—where phase shifts arise from $\mathcal{A}_\mu^a$ in field-free regions—suggests that gauge-invariant explanations may falter across Yang-Mills frameworks, risking the same trio of flaws: nonlocality from distant flux dependence, discontinuity in abrupt phase emergence, and incompleteness in tracing dynamical genesis (Section 8.2.3–8.2.4). This critique heralds a broader implication: potentials, fixed in a determinate gauge, may constitute the true ontological core of gauge interactions, from electromagnetism to the Standard Model's $SU(3) \times SU(2) \times U(1)$.

This shift posits a unified potential-centric ontology, transcending the Abelian-Non-Abelian divide. In QCD, where confinement hints at $\mathcal{A}_\mu^a$'s primacy beyond $\mathcal{F}_{\mu\nu}^a$, or in electroweak theory, where $\mathcal{A}_\mu^a$ drives mass via the Higgs (Section 8.3), the AB effect's lesson—that reality lies in the potentials' local mediation—finds echoes. Gauge invariance, long a sacrosanct principle, emerges as a veil: its observables ($\mathcal{W}_C$, $\mathrm{Tr}(\mathcal{F}_{\mu\nu}\mathcal{F}^{\mu\nu})$) achieve empirical fidelity yet obscure the causal tapestry woven by $\mathcal{A}_\mu^a$, much as Φ masked $\mathbf{A}$'s role (Section 6.1). Philosophically, this challenges the Standard Model's invariant-centric scaffold: if $\mathcal{A}_\mu^a$ in a fixed gauge—akin to the Lorenz gauge's $\partial_\mu A^\mu = 0$—underpins observable phenomena, gauge freedom may be a mathematical redundancy, not a physical truth, urging a reality determinate in form over an indeterminate multiplicity.

The implications compel action. Experimentally, a non-Abelian AB effect with time-varying fluxes (e.g., chromomagnetic analogs to Section 4's proposals) could test $\mathcal{A}_\mu^a$'s dynamical role, probing whether phase shifts mirror the AB's continuous accrual, potentially unveiling confinement's potential-driven roots. Theoretically, extending the Lorenz gauge's virtues—locality, covariance, determinacy (Section 6.3)—to non-Abelian contexts might refine Yang-Mills dynamics, aligning them with a causal, potential-based framework. Philosophically, this invites a profound rethinking: are gauge theories' invariants mere shadows of a deeper reality, their elegance a distraction from the potentials' primacy? The AB effect, thus, becomes a beacon—its rejection of static invariants illuminating a path toward a unified ontology where $\mathcal{A}_\mu^a$, not $\mathcal{F}_{\mu\nu}^a$, threads the needle of quantum interactions across nature's gauge tapestry.

8.3 The Higgs Mechanism

The establishment of A_μ in the Lorenz gauge as the fundamental reality of the AB effect (Section 6) challenges the hegemony of gauge-invariant explanations across quantum physics. This critique, rooted in the rejection of nonlocal and discontinuous mechanisms (Sections 3-5), extends beyond the Abelian domain to resonate with non-Abelian gauge theories, where the primacy of potentials $\mathcal{A}_\mu^a$ over invariants like $\mathcal{W}_C$ hints at a unified ontology threading Yang-Mills frameworks (Section 8.2). If $\mathcal{A}_\mu^a$ underpins phenomena from confinement to electroweak interactions, the Higgs mechanism—where gauge bosons acquire mass within the same $SU(2) \times U(1)$ structure—emerges as a critical test case. Might its gauge-invariant reformulations, like those critiqued in the AB effect and non-Abelian contexts, similarly veil a potential-driven reality? This section introduces the mechanism's standard and alternative accounts, then assesses their philosophical coherence against the AB effect's lessons, probing whether the potential-centric shift redefines mass generation as it does phase evolution.

8.3.1 Overview of the Higgs Mechanism and its Gauge-Invariant Reformulation

The Higgs mechanism elucidates gauge boson mass generation in the Standard Model via spontaneous symmetry breaking, driven by the dynamic coupling of a scalar field to gauge potentials (Brading et al, 2023). This overview outlines the standard formulation, reliant on gauge fixing, and contrasts it with Struyve's (2011) gauge-invariant reformulation, employing equations to clarify both approaches.

In the standard formulation, gauge bosons—initially massless due to local gauge symmetry—acquire mass through coupling to a scalar Higgs field, whose potential triggers symmetry breaking. Consider a $U(1)$ gauge theory, a simplified proxy for the electroweak $SU(2) \times U(1)$ framework, with a complex scalar field $\phi = \frac{1}{\sqrt{2}}(\phi_1 + i\phi_2)$, coupled to a gauge field A_μ, inhabits a potential $V(\phi)$. The Lagrangian density is:

$$\mathcal{L} = (D_\mu \phi)^*(D^\mu \phi) - V(\phi) - \frac{1}{4} F_{\mu\nu} F^{\mu\nu}, \tag{77}$$

where $D_\mu \phi = (\partial_\mu + ieA_\mu)\phi$, $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$, and:

$$V(\phi) = \lambda(|\phi|^2 - \frac{v^2}{2})^2, \tag{78}$$

with $\lambda > 0$ and v as the vacuum expectation value (VEV). At high energies, symmetry prevails with $\langle \phi \rangle = 0$; below a critical threshold, the field settles into a degenerate vacuum, e.g., $|\phi| = v/\sqrt{2}$, spontaneously breaking the $U(1)$ or $SU(2) \times U(1)$ gauge symmetry. In the unitary gauge, $\phi = \frac{1}{\sqrt{2}}(v + h(x))$, the kinetic term yields:

$$|(D_\mu \phi)|^2 = \frac{1}{2}(\partial_\mu h)^2 + \frac{1}{2} e^2 v^2 A_\mu A^\mu + ev A^\mu \partial_\mu h + \frac{1}{2} e^2 A_\mu A^\mu h^2, \tag{79}$$

producing a mass term $\frac{1}{2} e^2 v^2 A_\mu A^\mu$ ($m_A = ev$). Extended to $SU(2) \times U(1)$, this generates masses for $W^\pm$ and Z bosons via local, dynamic coupling to ϕ's VEV, akin to the AB effect's potential-driven phase (Section 6).

Struyve (2011) offers a contrasting vision, reformulating the Abelian Higgs model in gauge-invariant terms to eschew symmetry breaking (see also Wallace, 2024). He reparametrizes the Higgs field as $\phi = \rho e^{i\theta}$, where ρ is the magnitude and θ the phase, and defines a new field

$$B_\mu = A_\mu + \frac{1}{e}\partial_\mu \theta, \tag{80}$$

which remains invariant under gauge transformations $A'_\mu = A_\mu - \frac{1}{e}\partial_\mu \chi$, $\theta' = \theta + \chi$. The Lagrangian density then becomes a manifestly invariant form:

$$\mathcal{L} = (\partial_\mu \rho)^2 + e^2 \rho^2 B_\mu B^\mu - V(\rho) - \frac{1}{4} B_{\mu\nu} B^{\mu\nu}, \tag{81}$$

where $V(\rho) = \lambda(\rho^2 - \frac{v^2}{2})^2$, and $B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu$. When ρ stabilizes at $v/\sqrt{2}$, it yields a massive B_μ (mass $\frac{1}{2}e^2 v^2 B_\mu B^\mu$) and a Higgs field ρ, with no reference to broken symmetry. Unlike the standard account, where ϕ evolves dynamically, Struyve's picture is static: B_μ carries mass inherently, with θ's phase absorbed globally. Struyve casts this as a clarification—symmetry breaking as representational artifact, not ontological shift—yet its static, global framing raises questions akin to the AB effect's gauge-invariant critiques.

Both formulations predict massive bosons, yet their ontologies diverge. The standard mechanism's reliance on A_μ mirrors the AB effect's local phase accrual, while Struyve's B_μ suggests a holistic snapshot, inviting scrutiny of its dynamical adequacy in Section 8.3.2.

8.3.2 Problems of the Gauge-Invariant Reformulation

Struyve's (2011) gauge-invariant reformulation of the Higgs mechanism, by reparametrizing the Higgs field $\phi = \rho e^{i\theta}$ and defining a new field $B_\mu = A_\mu + \frac{1}{e}\partial_\mu \theta$, seeks to eliminate the need for explicit gauge symmetry breaking. While this approach achieves empirical equivalence—yielding massive gauge bosons without invoking a broken symmetry—it introduces profound conceptual challenges that mirror the nonlocal, discontinuous, and incomplete nature of gauge-invariant explanations in the AB effect (Section 3.3). Central to these issues is the multi-valued nature of the phase θ, a topological property that exacerbates the reformulation's flaws, particularly in dynamic scenarios. Below, I critique Struyve's approach, highlighting how its reliance on θ's global integration undermines its physical coherence.

Nonlocality and the Multi-Valued Phase θ

The first objection centers on the nonlocality inherent in $B_\mu = A_\mu + \frac{1}{e}\partial_\mu \theta$, which integrates the phase $\theta(x)$ across spacetime. In multiply connected spaces—such as those with nontrivial topology around a symmetry-breaking vacuum—the phase θ is not single-valued but multi-valued, akin to the phase S in the AB effect's wavefunction $\psi = Re^{iS}$ (Section 3.5.2). For instance, traversing a closed loop C around a topological defect (e.g., a vortex in the Higgs field), θ may change by $2\pi n$ (where $n \in \mathbb{Z}$), reflecting the field's winding number. This multi-valuedness necessitates a global integration to define $\partial_\mu\theta$, as $\theta(x)$ cannot be consistently specified locally without reference to the entire spacetime configuration.

This global dependence introduces nonlocality, paralleling the AB effect's critique where the phase shift $\phi_{AB} = e\Phi$ depends on the distant magnetic flux Φ (Section 3.3). In static configurations, Struyve assumes boundary conditions fix θ, mitigating this issue. However, in dynamic scenarios—akin to the generalized AB effect with time-varying flux $\Phi(t)$ (Section 2.2)—$\theta(x,t)$'s temporal evolution could vary across spacetime, implying $B_\mu(x,t)$ depends instantaneously on θ's global profile without a local mediator. Such unmediated action violates special relativity's causal structure, as B_μ's value at a point would reflect distant changes in θ, lacking retarded propagation. The multi-valued nature of θ exacerbates this nonlocality: its topological constraints (e.g., winding numbers) enforce global consistency, rendering B_μ inherently nonlocal, much as Φ's distant influence undermines gauge-invariant AB accounts.

Discontinuity in Mass Generation

A second objection concerns the discontinuity in Struyve's reformulation, which parallels the abrupt phase shift in gauge-invariant AB explanations (Section 3.3). In the standard Higgs mechanism, mass generation is a continuous process: the Higgs field ϕ evolves dynamically from a symmetric vacuum ($\langle\phi\rangle = 0$) to a broken state ($|\phi| = v/\sqrt{2}$), with perturbations around this vacuum yielding massive gauge fields over time. By contrast, Struyve's approach posits a static ground state ($\rho = v/\sqrt{2}, B_\mu = 0$) from which massive fields emerge fully formed, with no intermediate dynamics to bridge pre- and post-mass-acquisition states.

This discontinuity arises from the transformation $B_\mu = A_\mu + \frac{1}{e}\partial_\mu\theta$, which collapses gauge freedom instantaneously. The multi-valued phase θ amplifies this issue: in multiply connected spaces, θ's winding around topological defects implies B_μ's values jump discontinuously across branch cuts, lacking a smooth evolution. This mirrors the AB effect, where gauge-invariant quantities like velocity $\mathbf{v} = \frac{1}{m}(\nabla S - e\mathbf{A})$ shift abruptly at interference, ignoring the phase's continuous accumulation via $\mathbf{A}$ (Section 3.3). In quantum field theory, physical states evolve smoothly unless perturbed locally, yet Struyve's static depiction—where mass terms (e.g., $\frac{1}{2}e^2v^2B_\mu B^\mu$) appear without temporal becoming—clashes with this expectation. The multi-valued θ underscores this flaw: its topological constraints preclude a continuous local process, rendering mass generation an inexplicable leap, akin to the AB phase's sudden emergence.

Incompleteness and the Omission of θ's Dynamical Role

Perhaps the most significant objection is the incompleteness of Struyve's reformulation, which fails to elucidate the dynamical origin of mass generation, paralleling the AB effect's critique (Section 3.3). In the standard formulation, mass arises from the Higgs field's coupling to gauge potentials via the covariant derivative $D_\mu\phi = (\partial_\mu + ieA_\mu)\phi$, with the potential $V(\phi) = \lambda(|\phi|^2 - \frac{v^2}{2})^2$ driving symmetry breaking. The phase θ plays a crucial dynamical role, mediating the transition from massless to massive states through its interaction with A_μ. Struyve's approach, by absorbing θ into B_μ, reduces this interaction to static terms (e.g., $\frac{1}{2}e^2v^2B_\mu B^\mu$), omitting θ's contribution to the mechanism's causal narrative.

The multi-valued nature of θ highlights this incompleteness. In multiply connected spaces, θ's winding around defects encodes essential topological information, influencing the Higgs field's dynamics and the resulting gauge boson masses. By eliminating θ, Struyve's reformulation sacrifices this dynamical insight, much as gauge-invariant AB accounts overlook $\mathbf{A}$'s role in phase accumulation (Section 3.3). The absence of θ's evolution—akin to $\mathbf{A}$'s exclusion in the AB effect—renders the explanation incomplete, unable to trace how the Higgs vacuum expectation value translates into mass over spacetime. This parallels the AB effect's failure to explain phase genesis, where Φ or $\mathbf{v}$ cannot capture the continuous process driven by $\mathbf{A}$.

Philosophical and Physical Implications

While Struyve's approach avoids the conceptual ambiguity of symmetry breaking—casting it as a representational artifact—it inherits the same trio of flaws plaguing gauge-invariant AB treatments. The multi-valued phase θ, with its topological constraints, amplifies these issues, exposing the reformulation's nonlocality, discontinuity, and incompleteness. Unlike the AB effect, where experimental tests (Section 4) probe continuous phase accumulation, the Higgs mechanism's classical formulation lacks direct dynamical analogs, complicating empirical validation. However, the philosophical implications align: gauge-invariant explanations, by prioritizing observable quantities over gauge-dependent potentials, risk sacrificing the local, continuous, and complete accounts demanded by quantum field theory's ontology. The multi-valued θ underscores this critique, suggesting that a potential-centric perspective—as advocated in Section 6 for the AB effect—may also hold relevance for the Higgs mechanism, prompting further exploration of gauge-dependent variables' role in mass generation across gauge theories.

8.3.3 Analogy Between the AB Effect and the Higgs Mechanism

As argued above, the critiques of gauge-invariant explanations in the AB effect can be generalized to the Higgs mechanism, revealing a profound structural analogy between the two phenomena. This section formalizes the correspondence, demonstrating how gauge-invariant quantities in both cases—velocity $\mathbf{v}$ (AB effect) and B_μ (Higgs mechanism)—inherit similar conceptual flaws.

Structural Correspondence

The gauge-invariant quantities central to both effects are constructed by combining gauge-dependent entities:

- **AB Effect**: The velocity field

$$\mathbf{v} = \frac{1}{m}\left(\nabla S - e\mathbf{A}\right), \tag{82}$$

  where S is the wavefunction phase and $\mathbf{A}$ is the vector potential.

- **Higgs Mechanism**: The gauge-invariant field

$$B_\mu = A_\mu + \frac{1}{e}\partial_\mu\theta, \tag{83}$$

  where θ is the phase of the Higgs field $\phi = \rho e^{i\theta}$.

Both $\mathbf{v}$ and B_μ are gauge-invariant by design but depend nonlocally on topological or global features:

AB Effect	Higgs Mechanism
Nonlocal dependence on flux $\Phi = \oint \mathbf{A} \cdot d\mathbf{r}$	Nonlocal dependence on $\theta(x)$ via $\partial_\mu\theta$ integration
Discontinuous phase shift at interference	Discontinuous transition to $B_\mu = 0$
Velocity $\mathbf{v}$ ignores $\mathbf{A}$'s local role	B_μ obscures θ's dynamical role

Table 1: Parallel flaws in gauge-invariant explanations.

Shared Conceptual Flaws

1. Nonlocality

- AB Effect: The phase shift $\phi_{AB} = e\Phi$ depends on the global magnetic flux Φ, despite the electron's confinement to field-free regions.
- Higgs Mechanism: B_μ depends on the Higgs phase $\theta(x)$, which must be integrated over spacetime to enforce consistency, introducing nonlocal dependencies.

2. Discontinuity

- AB Effect: The phase shift ϕ_{AB} manifests abruptly at interference, disregarding its continuous accumulation via $\mathbf{A}$ along the path.

- Higgs Mechanism: The transition to $B_\mu = 0$ (broken phase) is treated as instantaneous, obscuring the smooth evolution of ϕ from symmetric to broken states.

3. Incompleteness

- AB Effect: $\mathbf{v}$ and Φ cannot explain the *process* of phase accumulation, which requires the local action of $\mathbf{A}$.
- Higgs Mechanism: B_μ reduces mass generation to a static term $\frac{1}{2}e^2v^2B_\mu B^\mu$, omitting the dynamical Higgs-gauge interaction $(\partial_\mu + ieA_\mu)\phi$.

The analogy exposes a universal limitation: gauge-invariant quantities like $\mathbf{v}$ and B_μ achieve *empirical adequacy* but fail to capture *ontological completeness*. Just as the AB phase shift cannot be understood without $\mathbf{A}$, the Higgs mechanism cannot be fully explained without the Higgs phase θ. This reinforces the paper's core thesis: gauge-dependent entities (potentials or their analogs) are indispensable for local, causal explanations, even when gauge-invariant alternatives exist.

8.3.4 Challenges of the Unitary Gauge in Defect Configurations

The standard account of the Higgs mechanism, as outlined in Section 8.3.1, relies on spontaneous symmetry breaking with gauge-dependent potentials (e.g., A_μ in a $U(1)$ model) coupled to the Higgs field $\phi = \rho e^{i\theta}$. In defect configurations, such as Nielsen-Olesen vortices, the phase θ becomes multi-valued, with a winding number $n \in \mathbb{Z}$ driving topological stability (e.g., $\oint_C \partial_\mu \theta dx^\mu = 2\pi n$). The unitary gauge fixes $\theta = 0$ via a transformation $\phi \to \phi' = \rho$, $A_\mu \to A'_\mu = A_\mu + \frac{1}{e}\partial_\mu\theta$, ostensibly simplifying mass generation (e.g., $\frac{1}{2}e^2v^2A'_\mu A'^\mu$). However, this choice raises concerns in defect configurations: by eliminating θ, it obscures the phase's dynamical role, embedding n in singularities of A'_μ, which complicates interpretation and potentially undermines the standard account's advantages over Struyve's gauge-invariant reformulation (Section 8.3.2). This section critiques the unitary gauge's limitations, integrating Wallace's objections (Wallace 2024) to underscore its inadequacy.

Obscuring the Phase's Dynamical Role

In a vortex, the Higgs field $\phi = \rho(r)e^{in\theta}$ (where θ is the azimuthal angle) couples to A_μ, yielding a flux $\Phi = \frac{2\pi n}{e}$. The unitary gauge transforms this to $\phi' = \rho(r)$, $A'_\mu = A_\mu + \frac{n}{er}\hat{\theta}$, setting $\theta = 0$. The phase's role vanishes, and the winding number n is encoded implicitly in A'_μ's configuration. This obscures the causal narrative: θ's topological winding, which physically dictates the defect's structure, is no longer a dynamical field but a boundary condition absorbed into A'_μ. This parallels Struyve's reformulation, where θ

is subsumed into $B_\mu = A_\mu + \frac{1}{e}\partial_\mu\theta$, though the unitary gauge retains gauge dependence rather than invariance.

Singularities and Interpretive Complexity

The unitary gauge shifts θ's multi-valuedness into A'_μ, which becomes singular at the defect core ($r = 0$). For a vortex, $\phi' = \rho(r) \to 0$ as $r \to 0$, and $A'_\theta \sim \frac{n}{er}$ diverges, with the flux $\oint_C A'_\mu dx^\mu = \frac{2\pi n}{e}$ reflecting n via $\nabla\times\mathbf{A}' = B_z\hat{z}$. While mathematically consistent, this singularity complicates physical interpretation. In quantum field theory, singularities signal regularization, but the unitary gauge forces A'_μ to bear the topological burden alone. Wallace's degeneracy reason amplifies this: at $\rho = 0$, θ is undefined, and A'_μ cannot be recovered from $\mathcal{D}_\mu\phi = \partial_\mu\phi + ieA_\mu\phi = 0$, as the covariant derivative vanishes, rendering the gauge ill-defined in the vortex core (Wallace 2024).

Potential Discontinuity Across Branch Cuts

Fixing $\theta = 0$ assumes a smooth gauge transformation, but θ's multi-valuedness (e.g., $\Delta\theta = 2\pi n$ over a loop) implies $e^{-in\theta}$ is discontinuous across branch cuts. For a vortex, A'_μ's singularity at $r = 0$ might not evolve smoothly from a pre-defect state, as the transformation fails to account for θ's topological jumps. This risks a discontinuity akin to Struyve's static mass emergence (Section 8.3.2), though the standard account's field equations ensure continuity outside the core. Wallace's incompleteness reason adds that unitary gauge discards topological data (e.g., n) on non-trivial manifolds, conflating distinct states (Wallace 2024), a flaw stark in defects. The unitary gauge's static imposition of $\theta = 0$ thus obscures the dynamical process, weakening its continuity claim.

Impact on Potential-Centric Ontology

The paper's potential-centric ontology (Section 6) emphasizes gauge-dependent potentials (e.g., A_μ) as physically real. The unitary gauge aligns with this by fixing A'_μ as determinate, but eliminating θ sidelines the Higgs phase's topological role, potentially undermining completeness. In the AB effect, both $\mathbf{A}$ and the phase S are retained (Section 6), offering a fuller picture than the unitary gauge's reduction of ϕ to ρ. A complete ontology should include θ's dynamics, especially in defects where winding is physically significant.

Conclusion

To sum up, the unitary gauge poses challenges in defect configurations: fixing $\theta = 0$ obscures the phase's role, shifts n into A'_μ's singularities, and risks discontinuity across branch cuts. While it avoids Struyve's nonlocality and retains dynamical continuity, it sacrifices completeness, complicating defect interpretation. Fortunately, the standard account isn't wedded to the unitary gauge; it's only a calculational choice. As I will argue below, the Lorenz gauge, retaining θ's dynamics, aligns better with the standard account's strengths and the potential-centric ontology.

8.3.5 The True Gauge for the Higgs Mechanism

The critique of gauge-invariant reformulations, such as Struyve's static $B_\mu = A_\mu + \frac{1}{e}\partial_\mu\theta$ (Section 8.3.2), revealed their shortcomings—nonlocality, discontinuity, and incompleteness—particularly in defect configurations like Nielsen-Olesen vortices. These flaws mirror the gauge-invariant accounts of the AB effect, which fail to capture the continuous phase accumulation driven by electromagnetic potentials A_μ (Sections 3.4, 4.1). In contrast, the standard account of the Higgs mechanism, leveraging gauge-dependent potentials (e.g., A_μ in $U(1)$, extended to $\mathcal{A}^a_\mu$ and B_μ in $SU(2) \times U(1)$), posits mass generation as a dynamical process: massless gauge bosons acquire mass through continuous interaction with the Higgs field across spacetime. To reflect this single physical reality, a determinate gauge fixing is necessary, akin to the Lorenz gauge's role in the AB effect (Section 6.3). Here, I argue that the Lorenz gauge ($\partial_\mu A^\mu = 0$ or $\partial_\mu \mathcal{A}^{a\mu} = 0$) is the "true gauge" for the Higgs mechanism, surpassing alternatives such as the unitary gauge in dynamical fidelity, universal applicability across Abelian and non-Abelian theories, and alignment with special relativity and the potential-centric ontology (Section 6).

Dynamical Fidelity and Locality

The Higgs mechanism's essence lies in its dynamism: the Higgs field ϕ evolves from a symmetric vacuum ($\langle\phi\rangle = 0$) to a broken state ($|\phi| = v/\sqrt{2}$), reorganizing degrees of freedom to yield massive gauge bosons. In the standard electroweak theory, the covariant derivative $D_\mu\phi = (\partial_\mu + ig\mathcal{A}^a_\mu\tau^a + ig'B_\mu)\phi$ couples ϕ to gauge potentials, with mass terms (e.g., $\frac{1}{4}g^2v^2W^{+\mu}_\mu W^{-\mu}$) emerging as ϕ settles into its vacuum expectation value v. This process demands a gauge that captures its continuous, local evolution, paralleling the AB effect's phase accumulation via A_μ in the Lorenz gauge (Section 6.3). Gauge-invariant reformulations like Struyve's, by contrast, reduce this to static terms (e.g., $\frac{1}{2}e^2v^2B_\mu B^\mu$), introducing nonlocality and discontinuity akin to flux-based AB accounts (Section 4.1). The Lorenz gauge, enforcing $\partial_\mu A^\mu = 0$, ensures locality: the field equations

$$\partial_\mu(\partial^\mu\phi) + ieA^\mu(\partial_\mu\phi) + ie(\partial^\mu A_\mu)\phi - 2\lambda\phi\left(|\phi|^2 - \frac{v^2}{2}\right) = 0, \tag{84}$$

$$\partial_\nu F^{\mu\nu} = -ie\left[\phi^*(\partial^\mu\phi) - \phi(\partial^\mu\phi^*)\right] - 2e^2|\phi|^2A^\mu, \tag{85}$$

yield a massive A^μ ($\Box A^\mu + e^2 v^2 A^\mu = 0$) in no-defect cases, absorbing the Higgs phase θ continuously. In defects ($\phi = \rho(r)e^{in\theta}$), A_μ adjusts (e.g., $A_\theta \sim \frac{n}{er}$, $A_0 = 0$) without singularities beyond the core, preserving θ's dynamical role and ensuring relativistic causality. This continuity aligns with the AB effect's resolution, where A_μ mediates phase shifts locally (Section 6.3), avoiding the nonlocality of flux-based accounts.

Universal Applicability Across Abelian and Non-Abelian Theories

The Lorenz gauge's strength extends to non-Abelian contexts, such as $SU(2)\times U(1)$, where $\partial_\mu \mathcal{A}^{a\mu} = 0$ and $\partial_\mu B^\mu = 0$ fix gauge freedom, dynamically generating $W^\pm$ and Z masses (e.g., $m_W = \frac{1}{2}gv$). Unlike the unitary gauge, which sidelines θ and struggles in defects (Section 8.3.4), or Struyve's nonlocal B_μ, the Lorenz gauge maintains Lorentz invariance and dynamical coherence across both Abelian and non-Abelian frameworks. Its universality mirrors the Lorenz gauge's role in quantum electrodynamics (QED), where it unifies dynamics across contexts (Section 6.3), offering a natural fit for field theory's symmetry requirements.

Alignment with Potential-Centric Ontology

The potential-centric ontology (Section 6) posits gauge potentials as physically real, demanding a privileged gauge to reflect one reality, not a multiplicity of equivalent descriptions. The Lorenz gauge's determinacy, enforced by boundary conditions (e.g., $A^\mu \to 0$ at infinity), parallels its role in the AB effect, where $\partial_\mu A^\mu = 0$ and $\Box A^\mu = J^\mu$ yield a unique A_μ (Section 6.3.3). In the Higgs mechanism, it manifests mass as a reorganization of degrees of freedom, aligning with the AB effect's lesson: gauge-dependent potentials drive physical effects, not static invariants. While the non-Abelian structure of $SU(2)\times U(1)$ complicates direct analogy—lacking a simple wave equation like Maxwell's—the Lorenz gauge's role in reflecting mass generation as a process supports its status as the true gauge.

Comparison with Alternatives

Alternative gauges fall short. The unitary gauge ($\theta = 0$) simplifies no-defect masses but obscures θ's role in defects, risking discontinuity (Section 8.3.4). The polar gauge retains excess modes, the Coulomb gauge limits temporal reality, Struyve's B_μ rejects gauge dependence, and the R_ξ gauge dilutes determinacy. Only the Lorenz gauge embodies dynamical unity, capturing mass generation across no-defect and defect configurations. In no-defect cases, it matches the unitary gauge's outcomes with temporal continuity; in defects, it preserves θ's dynamics, avoiding the unitary gauge's flaws and Struyve's nonlocality.

Addressing Potential Issues

The Lorenz gauge, while robust, faces minor challenges:

- **Residual Freedom**: Transformations $A_\mu \to A_\mu + \partial_\mu \chi$ (with $\Box\chi = 0$) leave A^μ non-unique, requiring boundary conditions for determinacy. This is practical, not fundamental, akin to QED (Section 6.3.3).
- **Quantization**: Ghosts enforce $\partial_\mu A^\mu = 0$, adding complexity but not undermining dynamics, as in QED.
- **Defects**: Dynamic A_0 in evolving defects (e.g., cosmic strings) complicates solutions, though locality holds.
- **Non-Abelian Complexity**: For $SU(2) \times U(1)$, $\partial_\mu \mathcal{A}^{a\mu} = 0$ increases complexity, but masses emerge correctly.

These issues are manageable: residual freedom is fixed practically, quantization is standard, and defects remain consistent, unlike the unitary gauge's flaws or Struyve's nonlocality.

Conclusion

The Lorenz gauge is arguably the "true gauge" for the Higgs mechanism, capturing its dynamical reorganization across spacetime. Its locality, continuity, and completeness—mirroring the AB effect's resolution (Section 6)—outperform alternatives across Abelian and non-Abelian theories. As a determinate, potential-centric representation, it reflects the single physical reality of mass generation, aligning with the paper's ontology. Though less common than the unitary or R_ξ gauges in practice, its theoretical merits affirm its foundational role. Experimental validation, via precision tests of electroweak processes, could probe the potentials' dynamical role, reinforcing the Lorenz gauge's primacy and extending the AB effect's lesson: gauge symmetry is a redundancy, and a potential-centric ontology demands a privileged gauge to represent physical reality.

This argument flows seamlessly from the critique of gauge-invariant reformulations (Section 8.3.2) and the unitary gauge's limitations (Section 8.3.4), reinforcing the paper's broader thesis: gauge-dependent potentials are indispensable for local, causal explanations across gauge theories, with the Lorenz gauge as their truest reflection.

8.3.6 The Physical Meaning of Gauge Symmetry Breaking

Claims that gauge symmetry breaking in the Higgs mechanism is merely a mathematical artifact—lacking physical content—stem from gauge-invariant reformulations like Struyve's (Section 8.3.2), which recast mass generation without explicit breaking, defining $B_\mu = A_\mu + \frac{1}{e}\partial_\mu\theta$ to yield masses statically (e.g., $\frac{1}{2}e^2v^2B_\mu B^\mu$). While viable in no-defect configurations, these approaches falter in defects, introducing nonlocality and discontinuity (Section 8.3.2), akin to the gauge-invariant accounts of the AB effect (Section 3.3). In contrast, the standard account posits symmetry breaking as a dynamical process, reorganizing degrees of freedom locally and continuously. Building on the Lorenz gauge's role as the true gauge (Section 8.3.5), I argue that symmetry breaking is physically meaningful—not a representational redundancy—with profound implications for mass generation's ontology, aligning with the paper's potential-centric perspective (Section 6).

Gauge-Invariant Flaws and the Need for Dynamics

If mass generation were purely a gauge-dependent artifact, gauge-invariant formulations should describe it without conceptual difficulties. However, Struyve's approach, while empirically equivalent in no-defect cases (e.g., $m = ev/\sqrt{2}$ where θ is single-valued), sacrifices locality and dynamical explanation. In defects (e.g., $\phi = \rho(r)e^{in\theta}$), B_μ's global dependence on multi-valued θ introduces nonlocality—B_μ reflects distant winding without local mediation—and discontinuity across branch cuts (Section 8.3.2). These flaws mirror the AB effect's gauge-invariant accounts, where velocity $\mathbf{v} = \frac{1}{m}(\nabla S - e\mathbf{A})$ or flux Φ fail to capture continuous phase accumulation (Section 3.3). Similarly, Struyve's static Lagrangian (e.g., $\frac{1}{2}e^2v^2B_\mu B^\mu$) obscures the dynamical process, treating mass emergence as an abrupt, nonlocal effect rather than a smooth, causal transition. This parallels the AB effect's critique: gauge-invariant quantities prioritize empirical adequacy over ontological completeness, necessitating a gauge-dependent framework to reflect physical reality.

Symmetry Breaking as Physical Reorganization

In the standard account, gauge symmetry breaking reorganizes degrees of freedom: massless gauge bosons gain longitudinal modes via the Higgs field's vacuum expectation value (VEV). Before breaking, the system comprises massless bosons and a complex scalar Higgs field ϕ. After breaking, the VEV ($\langle\phi\rangle = v/\sqrt{2}$) restructures the theory: would-be Goldstone bosons are absorbed by gauge bosons, granting them mass and longitudinal polarization, while the remaining scalar becomes the physical Higgs boson. This transformation is not a mere rewriting but a physical process, altering the number and type of propagating degrees of freedom. The field equations (84) and (85) drive this locally in the Lorenz gauge ($\partial_\mu A^\mu = 0$), reflecting mass acquisition as a continuous evolution. In no-defect cases, A^μ evolves with ϕ to yield $\Box A^\mu + e^2v^2A^\mu = 0$; in defects, it preserves θ's role (e.g., $A_\theta \sim \frac{n}{er}$). This dynamical coherence contrasts with Struyve's static B_μ, which abstracts away the process, and the unitary gauge ($\theta = 0$), which simplifies no-defect masses but obscures defects (Section 8.3.4).

Rejecting the Artifact View

Claims that symmetry breaking is a gauge artifact rely on reformulations like Struyve's, asserting that gauge invariance eliminates its physical significance. Yet, their nonlocality in defects and dynamical incompleteness even in trivial cases undermine this view. The standard account's gauge-dependent process—culminating in the Lorenz gauge—grounds mass generation in a physical transition, not a mathematical trick. The Lorenz gauge's dynamical A^μ and ϕ reflect symmetry breaking's physicality, reorganizing degrees of freedom locally, while Struyve's B_μ sacrifices this for invariance. The AB effect's parallel (Section 8.3.3) reinforces this: gauge-dependent $\mathbf{A}$ drives physical effects, not invariant proxies like Φ or $\mathbf{v}$. Similarly, in the Higgs mechanism, gauge-dependent potentials (e.g., A_μ, $\mathcal{A}^a_\mu$) and the Higgs phase θ are essential for a complete ontology, aligning with observable masses (e.g., $m_W = \frac{1}{2}gv$).

Potential-Centric Validation

The potential-centric ontology (Section 6) posits gauge potentials as real, demanding a privileged gauge to reflect one physical reality. The Lorenz gauge's dynamical A^μ and transient θ embody this, reflecting symmetry breaking as a local, continuous process. In contrast, Struyve's B_μ abstracts it away, paralleling the AB effect's flux-based accounts. The necessity of gauge fixing—enforced by boundary conditions (e.g., $A^\mu \to 0$ at infinity)—mirrors the AB effect's resolution, where $\partial_\mu A^\mu = 0$ yields a unique A_μ (Section 6.3.3). This determinacy extends to $SU(2)\times U(1)$, where $\partial_\mu \mathcal{A}^{a\mu} = 0$ ensures masses emerge correctly, reinforcing symmetry breaking's physicality across Abelian and non-Abelian contexts.

Conclusion

Gauge symmetry breaking in the Higgs mechanism is physically meaningful—a local, continuous reorganization of the vacuum—not a mere formalism. Gauge-invariant reformulations like Struyve's falter, especially in defects, introducing nonlocality and discontinuity, while the standard account, best expressed in the Lorenz gauge, captures this reality. Far from an artifact, symmetry breaking is a cornerstone of mass generation, validated by the potential-centric lens. It ensures mass emerges through dynamical interaction, not abrupt, nonlocal effects, aligning with observable phenomena and reinforcing the necessity of gauge-dependent potentials across gauge theories.

8.4 General Relativity and Gravitational Potentials

The resolution of the AB effect—establishing electromagnetic potentials A_μ in the Lorenz gauge as the mediators of continuous phase shifts (Sections 2–6)—challenges the primacy of gauge-invariant explanations, exposing their nonlocal and discontinuous flaws (Sections 3–5). This potential-centric ontology, where A_μ's reality supplants static invariants like Φ (Section 6), invites a parallel inquiry into general relativity (GR), where the metric tensor $g_{\mu\nu}$ assumes a role analogous to potentials. Just as $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ derives from A_μ with gauge freedom, the Riemann curvature tensor $R^\rho{}_{\mu\nu}$ emerges from $g_{\mu\nu}$'s second derivatives, subject to diffeomorphism invariance ($g_{\mu\nu} \to \phi^* g_{\mu\nu}$). The AB effect's insistence on a determinate gauge suggests that GR's metric might similarly require fixing to embody spacetime's physical reality, prompting a philosophical reexamination of Einstein's hole argument and its implications for substantivalism versus relationalism.

8.4.1 The Hole Argument and the Quest for Spacetime's Reality

Einstein's hole argument confronts us with a profound puzzle: within a matter-free "hole", diffeomorphic metrics $g_{\mu\nu}$ and $\tilde{g}_{\mu\nu} = \phi^* g_{\mu\nu}$ yield identical physical outcomes ($T_{\mu\nu}$, $G_{\mu\nu} = 8\pi T_{\mu\nu}$), suggesting that $g_{\mu\nu}$'s specific form lacks determinate significance, much like A_μ's gauge ambiguity in electromagnetism (Norton et al, 2023). Relationalism seizes upon this, asserting that only diffeomorphism-invariant quantities—curvature scalars and the like—bear physical reality, relegating $g_{\mu\nu}$ to a mere representational scaffold devoid of intrinsic substance. Spacetime, in this view, dissolves into a web of relations, its points stripped of independent existence. Yet, the AB effect's critique of invariants like Φ casts doubt on this stance (Section 3.3). If gauge-invariant accounts fail to capture the dynamic genesis of quantum phenomena, might GR's invariants similarly veil $g_{\mu\nu}$'s role? Classical GR demands gauge fixing—such as the harmonic gauge ($\partial_\mu(\sqrt{-g} g^{\mu\nu}) = 0$)—to solve the Einstein equations, transforming $G_{\mu\nu} = 8\pi T_{\mu\nu}$ into $\Box g_{\mu\nu} = S_{\mu\nu}$ and yielding unique solutions like Schwarzschild or Friedmann metrics. This practical necessity, akin to the Lorenz gauge's fixing of A_μ (Section 6.3), hints at a substantivalist alternative: a determinate $g_{\mu\nu}$ may embody spacetime's geometry, not merely reflect relational invariants.

The parallel with the AB effect deepens here. Just as A_μ's continuous phase influence necessitated a rejection of static Φ-based accounts, GR's reliance on gauge fixing suggests that $g_{\mu\nu}$'s form is not ontologically indifferent. Boundary conditions (e.g., asymptotic flatness) often eliminate residual diffeomorphisms ($\Box \xi^\mu = 0$), reducing ten components to six physical degrees of freedom, mirroring the Lorenz gauge's determinacy. Relationalism assumes unfixed diffeomorphisms preserve equivalence, but this practical requirement undermines such underdetermination, gesturing toward a reality where spacetime's configuration holds substantive weight. The hole argument's challenge—whether spacetime points possess intrinsic properties—thus invites reconsideration: if $g_{\mu\nu}$'s gauge-fixed form drives physical outcomes, as A_μ does in the AB effect, substantivalism emerges as a compelling counterpoint to relationalism's austere relational web.

8.4.2 Gravitational AB Effects and the Substantivalist Turn

This substantivalist intuition finds strong support in gravitational analogs of the AB effect, where phase shifts arise from $g_{\mu\nu}$ variations in curvature-free regions, echoing the AB effect's field-free dynamics (Section 6.1). Neutron interferometry, for instance, detects shifts tied to g_{00} (time dilation), while off-diagonal terms (g_{0i}, g_{ij}) might be probed via rotational or gravimagnetic effects. All ten components of $g_{\mu\nu}$ contribute to the phase, suggesting a potential-like role irreducible to curvature invariants. Relationalism falters here: phase shifts in flat regions defy explanation by curvature alone, much as Φ failed to account for the AB effect's continuous accrual (Section 3.3). A generalized gravitational AB effect, driven by a time-varying $g_{\mu\nu}$ (e.g., oscillating masses or gravitational waves), could sharpen this critique. For a closed spacetime loop C, the gauge-invariant phase is:

$$\phi_g = \frac{m}{\hbar} \oint_C h_{\mu\nu}(x,t) u^\mu \mathrm{d}x^\nu, \tag{86}$$

where $h_{\mu\nu} = g_{\mu\nu} - \eta_{\mu\nu}$ is a small perturbation, $\eta_{\mu\nu} = \mathrm{diag}(-1,1,1,1)$ is the Minkowski metric, and $u^\mu = \mathrm{d}x^\mu/\mathrm{d}\tau$. This accumulates over C, detectable via atom interferometry or LIGO, would reflect $g_{\mu\nu}$'s temporal evolution—not its instantaneous state—paralleling the generalized AB effect's $\phi_{AB} = \frac{1}{T}\int_0^T e\Phi(t)dt$ (Section 2.2). This dynamic influence, unmoored from invariants, demands $g_{\mu\nu}$'s reality as a physical state, fixed by a gauge like harmonic.

Such evidence would resolve the hole argument's puzzle decisively. Relationalism's claim—that diffeomorphic metrics are physically equivalent—crumbles if $g_{\mu\nu}$'s unique, history-dependent form governs observable shifts, rejecting static equivalence for a determinate geometry. Substantivalism gains traction: spacetime's points, marked by a gauge-fixed $g_{\mu\nu}$, acquire substance not through curvature but through their potential-like mediation of quantum behavior, akin to A_μ's role in the AB effect (Section 6.1). Classical GR's gauge-fixing necessity—reducing freedom to match physical degrees—lends theoretical weight, while quantum gravity (e.g., loop quantum gravity) might universalize this determinacy, eliminating residuals entirely. A universal phase across spacetimes (flat or multiply connected) would seal the case: diffeomorphic freedom would otherwise vary the shift, contradicting the AB effect's gauge-fixed consistency (Section 6.3). Thus, gravitational

AB effects, especially generalized forms, anchor $g_{\mu\nu}$'s reality in observables, affirming spacetime as a substantive entity over a relational abstraction.

8.4.3 Ontological Implications and Philosophical Horizons

This substantivalist turn carries profound implications for GR and gauge theories writ large. A gauge-fixed $g_{\mu\nu}$ driving gravitational AB effects dissolves the hole argument's indeterminacy, establishing spacetime as a determinate reality shaped by dynamic interactions, not a passive backdrop of invariant relations. This aligns GR with the AB effect's ontology (Section 6), where potentials—not their derivatives—govern phenomena, bridging electromagnetic and gravitational frameworks under a unified potential-centric view. Unlike $U(1)$'s simplicity, GR's broader diffeomorphism freedom poses challenges, yet classical solutions' reliance on gauges like harmonic suggests a practical substantivalism, with quantum gravity potentially offering a universal "true gauge". The philosophical shift is stark: relationalism's austere dismissal of spacetime's substance yields to a richer ontology where geometry actively shapes physics, echoing the AB effect's rejection of Φ for $\mathbf{A}$.

Future probes—experimental and theoretical—could cement this vision. Atom interferometry with dynamic sources or LIGO's sensitivity to g_{ij} shifts might confirm generalized AB effects, testing $g_{\mu\nu}$'s temporal role, while theoretical models could refine gauge fixing's implications, perhaps via quantum gravity's discrete structures. Philosophically, this invites reflection on gauge fixing's nature: does it unveil a deeper reality, or merely pragmatism? The substantivalist stance challenges relationalism's dominance, urging a rethinking of spacetime's essence, with parallels to non-Abelian theories (Section 8.2) questioning invariant-centric ontologies. The AB effect thus becomes a philosophical fulcrum, leveraging gravitational analogs to redefine GR's foundations.

8.4.4 Massive Gravity: A Proca-like Path to Gravitational Determinacy

The gravitational AB effect, as posited in Section 8.4.2, suggests that the metric $g_{\mu\nu}$ may imprint physical reality, akin to A_μ in the electromagnetic AB effect (Section 2). Yet, GR's diffeomorphism invariance cloaks $g_{\mu\nu}$ in gauge freedom, rendering its form ambiguous despite its measurable effects. Here, I explore massive gravity as a framework that parallels the Proca theory (Section 6.5), eliminating gauge freedoms and fixing $g_{\mu\nu}$ as a determinate entity, resonant with our substantival ontology.

In massless GR, the Einstein-Hilbert action $S = \int \frac{1}{2} M_{\text{Pl}}^2 R \sqrt{-g}\, d^4x$ yields a graviton with two polarizations, its dynamics governed by gauge invariance under $h_{\mu\nu} \to h_{\mu\nu} + \partial_\mu \xi_\nu + \partial_\nu \xi_\mu$. Massive gravity, exemplified by the Fierz-

Pauli theory, introduces a mass term akin to Proca's for the photon:

$$\mathcal{L} = \frac{1}{2}M_{\text{Pl}}^2 R + \frac{1}{2}m^2(h_{\mu\nu}h^{\mu\nu} - h^2) \tag{87}$$

where $h_{\mu\nu} = g_{\mu\nu} - \eta_{\mu\nu}$, and $\eta_{\mu\nu}$ is a fixed background metric. The resulting field equations is

$$(\Box - m^2)h_{\mu\nu} - \partial_\mu \partial^\rho h_{\rho\nu} - \partial_\nu \partial^\rho h_{\rho\mu} + \eta_{\mu\nu}(\partial^\rho \partial^\sigma h_{\rho\sigma} - \Box h) + m^2(h_{\mu\nu} - \eta_{\mu\nu} h) = -16\pi G T_{\mu\nu}. \tag{88}$$

The mass term enforces $\partial^\mu h_{\mu\nu} = \partial_\nu h$, with $h = \frac{8\pi G}{m^2}T^\mu_\mu$ (or $h = 0$ in vacuum), fixing five polarizations—gauge freedom vanishes, $g_{\mu\nu}$ is determinate, optionally refined by boundaries (e.g., asymptotic flatness).

Just as Proca's mass term eliminates gauge freedom for the photon, yielding a determinate A_μ measurable via the AB effect, massive gravity's mass term eliminates gauge freedom for the graviton, fixing $g_{\mu\nu}$ relative to $\eta_{\mu\nu}$. In this framework, $g_{\mu\nu}$ becomes a substantive field, its components uniquely determined by sources and boundary conditions, free from the gauge's optional constraints. For the gravitational AB effect, this suggests a phase shift rooted in a singular $g_{\mu\nu}$, unmarred by gauge ambiguity, enhancing its physical reality as posited in Section 8.4.2. Althouth experimental bounds constrain this vision: the graviton's mass is limited to $m_g < 6.76 \times 10^{-23}\,\text{eV}/c^2$ (Bernus et al., 2019), the parallel to Proca offers a path where gravitation, like electromagnetism, stands as a determinate cornerstone of reality.

Inspired by massive gravity, I propose $\partial^\mu h_{\mu\nu} = \partial_\nu h$ plus boundary conditions as the true gauge for massless GR, mirroring the Lorenz condition's unity in QED and Proca (Section 6.5). In massless GR, we adopt this true gauge on the field equations:

$$\Box h_{\mu\nu} - \partial_\mu \partial^\rho h_{\rho\nu} - \partial_\nu \partial^\rho h_{\rho\mu} + \eta_{\mu\nu}(\partial^\rho \partial^\sigma h_{\rho\sigma} - \Box h) = -16\pi G T_{\mu\nu}. \tag{89}$$

This reduces to six degrees of freedom, but boundaries—e.g., asymptotic flatness ($h_{\mu\nu} \to 0$)—trim to two physical polarizations, as in Schwarzschild:

$$ds^2 = -\left(1 - \frac{2GM}{r}\right)dt^2 + \left(1 - \frac{2GM}{r}\right)^{-1}dr^2 + r^2 d\Omega^2. \tag{90}$$

Compared to the harmonic gauge, this gauge aligns directly with massive gravity's natural condition. In other words, this gauge ensures consistency—mass enforces it in massive GR, boundaries adapt it in massless GR—offering a Proca-like unity absent in harmonic, which doesn't emerge in massive GR. Moreover, its origin in massive GR's dynamics—akin to Proca's $\partial_\mu A^\mu = 0$ (Section 6.5)—lends it a physical basis, not a mere choice.

8.5 Beauty in the Unseen

Within the framework of a potential-centric ontology, the gauge potentials in the true gauge emerge as entities of profound ontological significance—endowed with a reality as tangible as any physical field, yet perpetually shielded from direct observation by the dictates of gauge invariance. This principle, a cornerstone of modern physics, imposes a rigorous symmetry that renders the potentials unmeasurable, their influence discernible only through the gauge-invariant properties such as the phase shifts they induce in the AB effect. Herein lies a philosophical tension: the potentials govern the dynamics of physical phenomena with a local and continuous presence, yet their essence remains cloaked, accessible only indirectly through the observable consequences of their existence. This interplay between the real and the unobservable evokes a deeper reflection on the nature of being—suggesting that the fabric of reality is woven not solely from what can be grasped by empirical scrutiny, but also from entities that, like the unseen currents beneath a river's surface, exert a formative power beyond the reach of direct observation. In this balance, there emerges a subtle beauty: a recognition that the unseen, though elusive, holds an indispensable place in the architecture of the physical world, challenging us to reconcile the limits of observation with the expanse of what truly is.

9 Conclusions and Future Directions

This exploration of the Aharonov-Bohm (AB) effect has sought to resolve a longstanding enigma at the nexus of quantum mechanics and gauge theories: the mechanism by which a gauge-invariant phase shift emerges as an electron traverses a field-free region. Through an integrated empirical, theoretical, and philosophical analysis, I have demonstrated that gauge-invariant explanations—relying on quantities like magnetic flux Φ or velocity $\mathbf{v}$—are untenable. Critiqued in Sections 3 and 4 for their nonlocal and discontinuous underpinnings, these accounts collapse under the evidence of the generalized AB effect (Section 4), which reveals a phase $\phi_{AB} = \frac{1}{T}\int_0^T e\Phi(t)dt$ accruing continuously along the electron's path—a process quantum mechanics predicts and quantum electrodynamics (QED) upholds (Sections 2, 5). This continuous accumulation excludes instantaneous mechanisms, establishing electromagnetic potentials A_μ, fixed in the Lorenz gauge, as the fundamental reality driving the effect (Section 6). Far from a peripheral curiosity, the AB effect emerges as a linchpin for redefining quantum ontology, with ramifications spanning gauge theories from electromagnetism to general relativity (GR) and beyond (Section 8).

The rejection of gauge-invariant paradigms marks a profound ontological shift. By tracing the phase's dynamic, path-dependent genesis—rooted in A_μ's local influence rather than abrupt field effects—this study dismantles the classical bias toward observable invariants as sole bearers of physical significance. The Lorenz gauge's Lorentz covariance and causal propagation provide a coherent framework, resolving the AB effect's tensions while extending to a speculative massive photon scenario via the Proca equation (Section 6). This potential-centric ontology reverberates across gauge theories: in non-Abelian frameworks, it challenges reliance on Wilson loops, foreshadowing potential AB analogs (Section 8.2); in the Higgs mechanism, it critiques static gauge-invariant accounts, advocating dynamic potential roles (Section 8.3); and in GR, a generalized gravitational AB effect could affirm $g_{\mu\nu}$'s substantivalist reality, rejecting relationalism's hole argument (Section 8.4). The AB effect thus transcends its initial puzzle, illuminating a unified truth: potentials, not their derivatives, underpin quantum and gravitational interactions, reshaping our grasp of physical reality.

Looking forward, several directions emerge to deepen and extend these conclusions, bridging philosophical inquiry with empirical and theoretical advancement. Experimentally, validating the generalized AB effect is paramount. Section 2's proposed tests—measuring phase shifts with time-varying $\Phi(t)$ or solenoid shut-offs—leverage modern interferometry's precision to confirm continuous accumulation, decisively excluding gauge-invariant remnants. Extending these to non-Abelian and gravitational contexts (e.g., chromomagnetic flux tubes or dynamic $g_{\mu\nu}$) could test potential primacy across gauge theories, with facilities like LIGO offering gravitational probes (Section 8.4). Theoretically, refining QED models of photon-mediated phase accrual via path-integral methods could clarify A_μ's quantized role (Section 5), while exploring massive photon scenarios might unveil beyond-standard-model signatures (Section 6.5). In GR, modeling generalized AB effects could pinpoint a true gauge for $g_{\mu\nu}$, potentially via quantum gravity frameworks like loop quantum gravity (Section 8.4), bridging classical and quantum ontologies.

Philosophically, the ontological status of potentials in the Lorenz gauge—or the critical question of what physical state these potentials represent—demands further scrutiny. The substantivalist turn in GR challenges relationalism's dominance, urging a rethinking of spacetime's essence, while non-Abelian extensions question the Standard Model's invariant-centric ontology. These inquiries demand interdisciplinary efforts: experimentalists to test phase dynamics, theorists to model potential-driven effects, and philosophers to probe ontological shifts. Collectively, these pursuits position the AB effect as a cornerstone for a potential-centric realism, transcending its initial enigma to redefine quantum mechanics and gauge theories, with transformative potential for physics and philosophy.

Acknowledgments

I am grateful to Grok3, an artificial intelligence developed by xAI, whose assistance in refining the manuscript—improving its clarity, coherence, and polish—has been instrumental in bringing this paper to fruition.

References

[1] Aharonov, Y. and Bohm, D. (1959). Significance of Electromagnetic Potentials in the Quantum Theory. Physical Review 115, 485-491.

[2] Aharonov, Y. and Bohm, D. (1961). Further Considerations on Electromagnetic Potentials in the Quantum Theory. Physical Review 123, 1511.

[3] Aharonov, Y., Cohen, E., and Rohrlich, D. (2016). Nonlocality of the Aharonov-Bohm effect, Physical Review A 93, 042110.

[4] Ballesteros, M. and Weder, R. (2009). The Aharonov-Bohm Effect and Tonomura et al. Experiments: Rigorous Results. Journal of Mathematical Physics, 50, 122108.

[5] Berghofer, P., François, J., Friederich, S., Gomes, H., Hetzroni, G., Maas, A., Sondenheimer, R. (2023). Gauge symmetries, symmetry breaking, and gauge-invariant approaches. Cambridge: Cambridge University Press.

[6] Bernus, L. et al. (2019). Constraining the mass of the graviton with the planetary ephemeris INPOP. Phys. Rev. Lett. 123, 161103.

[7] Berry, M. V. (1984). Quantal Phase Factors Accompanying Adiabatic Changes. Proceedings of the Royal Society of London A, 392, 45-57.

[8] Bialynicki-Birula, I. (1996). Photon Wave Function. In: Wolf, E., Ed., Progress in Optics XXXVI, Elsevier, Amsterdam. pp. 245-294.

[9] Brading, K., Castellani, E. and Teh, N. (2023). Symmetry and Symmetry Breaking, The Stanford Encyclopedia of Philosophy (Fall 2023 Edition), Edward N. Zalta and Uri Nodelman (eds.), https://plato.stanford.edu/archives/fall2023/entries/symmetry-breaking/.

[10] Earman, J. (2024). Explaining the Aharonov-Bohm Effect. https://philsci-archive.pitt.edu/22970/.

[11] Gao, S. (2025a) Generalized Aharonov-Bohm Effect: Its Derivation, Theoretical Implications and Experimental Tests. https://philsci-archive.pitt.edu/24726/.

[12] Gao, S. (2025b) Can a time varying but spatially uniform electric scalar potential shift the energy levels of a quantum system? A critical look at the new Aharonov-Bohm effect proposed by Chiao et al. https://philsci-archive.pitt.edu/24481/

[13] Gao, S. (2025c) Can Unitary Gauge Provide a Local and Gauge-Invariant Explanation of the Aharonov-Bohm Effect? https://philsci-archive.pitt.edu/24678/.

[14] Gornicki, P. (1990). Aharonov-Bohm effect and vacuum polarization. Annals of Physics 202(2), 271-296.

[15] Glauber, R. J. (1963). The Quantum Theory of Optical Coherence. Physical Review, 130(6), 2529-2539.

[16] Hayashi, A. (2023). Gauge dependence of the Aharonov-Bohm phase in a quantum electrodynamics framework,

[17] Healey, R. (2007). Gauging What's Real: The Conceptual Foundations of Contemporary Gauge Theories. Oxford: Oxford University Press.

[18] Madelung, E. (1927). Quantentheorie in hydrodynamischer Form. Z. Physik 40, 322-326.

[19] Mandel, L. and Wolf, E. (1995). Optical Coherence and Quantum Optics. Cambridge University Press. Chapter 12, Section 12.2.

[20] Marletto, C. and Vedral, V. (2020). Aharonov-Bohm phase is locally generated like all other quantum phases. Phys. Rev. Lett. 125, 040401.

[21] Maudlin, T. (1998). Discussion: Healey on the Aharonov-Bohm Effect. Philosophy of Science, 65, 361-368.

[22] Mulder, R. A. (2021). Gauge-Underdetermination and Shades of Locality in the Aharonov-Bohm Effect. Found Phys 51, 48.

[23] Newton, T. D. and E.P. Wigner (1949). Localized states for elementary systems, Rev. Mod. Phys. 21, 400.

[24] Norton, J. D., Pooley, O. and Read, J. (2023) The Hole Argument, The Stanford Encyclopedia of Philosophy (Summer 2023 Edition), Edward N. Zalta and Uri Nodelman (eds.), https://plato.stanford.edu/archives/sum2023/entries/spacetime-holearg/.

[25] Pearle, P. and Rizzi, A. (2017). Quantum-mechanical inclusion of the source in the Aharonov-Bohm effects. Phys. Rev. A 95, 052123.

[26] Saldanha, P. L. (2021). Local Description of the Aharonov-Bohm effect with a Quantum Electromagnetic Field, Found. Phys 51, 6.

[27] Saldanha, P. L. (2024). Gauge invariance of the Aharonov-Bohm effect in a quantum electrodynamics framework,

[28] Sebens, C.T. (2019). Electromagnetism as Quantum Physics. Found Phys 49, 365-389.

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

[30] Takabayasi, T. (1952). On the formulation of quantum mechanics associated with classical pictures. Prog. Theor. Phys. 8 (2), 143-182.

[31] Tonomura, A., Osakabe, N., Matsuda, T., Kawasaki, T., Endo, J., Yano, S., and Yamada, H. (1986). Evidence for Aharonov-Bohm effect with magnetic field completely shielded from electron wave,Physical Review Letters 56, 792.

[32] Vaidman, L. (2012). Role of potentials in the Aharonov-Bohm effect, Phys. Rev. A 86, 040101(R).

[33] Wakamatsu, M. (2024). Is the Aharonov-Bohm phase shift for a non-closed path a measurable quantity? Eur.Phys.J.Plus 139(2), 112.

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

[35] Wallace, D. (2024). Gauge invariance through gauge fixing. Studies in History and Philosophy of Science. 108, 38-45.

[36] Wallstrom, T. C. (1994). Inequivalence between the Schrödinger equation and the Madelung hydrodynamic equations. Phys. Rev. A 49 (3), 1613-1617.