UNIVERSITY OF CALIFORNIA,
IRVINE

Symmetry and Equivalence in The Constrained Hamiltonian Formalism

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Philosophy

by

Clara Bradley

Dissertation Committee:
Chancellor's Professor James Owen Weatherall, Chair
Chancellor's Professor Jeffrey A. Barrett
Professor JB Manchak

2024

DEDICATION

To my parents Richard and Shura and my husband Carver, for always being my rocks.

TABLE OF CONTENTS

LIST OF FIGURES

Page

ACKNOWLEDGEMENTS

My first thanks go to my supervisor, Jim. From the day I met Jim in 2016, he has inspired me, supported me, and advocated for me beyond compare. Jim has taught me to be a clearer writer, speaker, and thinker, to believe in myself, and to have fun while doing it. I am forever grateful for Jim's generosity and assurance, and I feel very lucky to have had him as my supervisor.

Thank you to my other committee members Jeff Barrett and JB Manchak. Jeff and JB's feedback has helped me to better motivate and present the work in this dissertation, and I have learnt a lot from them on how to do technical philosophy precisely and clearly.

Thank you to all the members of PPRG over the last 5 years. PPRG has been a source of accountability, but also of friendship. Special thanks to those who have given feedback on drafts of papers and practice talks, including Jingyi Wu, Helen Meskhize, Ellen Shi, Aidan Carter, and India Bhalla-Ladd.

Thank you to Thomas Barrett for inviting me to Santa Barbara to talk about some of the ideas in this dissertation. Thank you to David Wallace for several conversations and for sending me drafts of papers that have been extremely helpful. Thank you to Neil Dewar for inviting me to be part of a PSA symposium that led to Chapter 3 of this dissertation. Thank you to James Ladyman, Karim Thébault, Henrique Gomes, Sean Gryb, Jeremy Butterfield, and Brian Pitts for conversations and for support.

Thank you to my parents and my brother. I am indebted to your love, kindness, and commitment. Dad, thank you for all the walks where we chatted about philosophy, for all the times you have read my work or given me advice, and for keeping my expectations realistic. Despite your best attempts, I am a philosopher because of you. Mum, thank you for always picking up the phone with the same excitement every time; speaking to you every few days has made being so far from home bearable. Adrian, thank you for being the type of brother who keeps me in check but who also keeps an eye out for me.

Thank you to my extended family, both on the Bradley side and the Sorensen side. I know that you all wonder to some extent why I am doing this, and I am grateful that despite this, you have all supported me and my career choice.

Thank you to all of my amazing, supportive friends who remind me why I pursued a PhD as well as what is important outside of academia, especially Ellen Muller, Nick Clanchy, Nick Daultry-Ball, Caroline Jones, Ben Genta, Margaret Farrell, and Rebecca Korf.

Thank you to my husband, Carver. Meeting you and getting married to you has made writing my PhD feel like part of a romantic movie. You have made me a better person, and a happier one. Thank you for your unconditional love and support and for all the sacrifices you make to let me pursue my goals. Going through life with you is greater than anything else.

This dissertation is partly made possible through generous support from the Social Science Merit Fellowship, the Provost PhD Fellowship, and the Graduate Dean's Recruitment Fellowship. Materials from this dissertation were presented at PSA 2022 in Pittsburgh, Foundations of Physics 2021 in Paris, Foundations of Physics 2023 in Bristol, New Directions in the Foundations of Physics 2024 in Slovenia, the Foundations of Physics @ Harvard seminar, the University of Bristol Philosophy of Science Seminar, the Quantum Spacetime in the Cosmos Conference at Perimeter Institute, and the Symmetry and Structure Workshop in Irvine.

Chapter Three is published Open Access under license CC BY 4.0 in *Philosophy of Science*: 'The Representational Role of Sophisticated Theories' by Clara Bradley, 2023. The published version is accessible at https://doi.org/10.1017/psa.2023.131. Minor changes were made.

VITA

Clara Bradley

Ph.D in Philosophy **2024**
University of California, Irvine *Irvine, California*

M.Sci. in Physics and Philosophy **2019**
University of Bristol *Bristol, United Kingdom*

PUBLICATIONS

The Representational Role of Sophisticated Theories **2023**
Philosophy of Science

Mathematical Responses to the Hole Argument: Then and Now (with James Owen Weatherall) **2022**
Philosophy of Science

The Non-Equivalence of Einstein and Lorentz **2021**
The British Journal for the Philosophy of Science

On Representational Redundacy, Surplus Structure, and the Hole Argument (with James Owen Weatherall) **2020**
Foundations of Physics

ABSTRACT OF THE DISSERTATION

Symmetry and Equivalence in The Constrained Hamiltonian Formalism

By

Clara Bradley

Doctor of Philosophy in Philosophy

University of California, Irvine, 2024

Chancellor's Professor James Owen Weatherall, Chair

How do we characterize the symmetries of a theory? How should we respond to the presence of 'excess structure' in a theory? When are two theories equivalent? This dissertation is an exploration of these questions in the context of a particular formulation of classical theories known as the constrained Hamiltonian formalism.

A theme running through the dissertation is that progress can be made on these questions by making precise the geometric structure of a constrained Hamiltonian theory. I argue that this geometric structure can (1) be used to resolve a debate about how to correctly characterize the relationship between constraints and gauge transformations, (2) shed light on the relationship between the constrained Hamiltonian formalism and the Lagrangian formalism, and (3) provide an avenue for formulating a constrained Hamiltonian theory that resolves an apparent tension between gauge variables being 'excess structure' and playing an ineliminable role.

Introduction

The topic of this dissertation is the mathematical foundations of the constrained Hamiltonian formalism and its connection to philosophical questions concerning theoretical equivalence, excess structure, and the interpretation of symmetries. Most philosophical papers on the constrained Hamiltonian formalism are centered around a puzzle known as 'The Problem of Time'.[1] The Problem of Time is multi-faced, but the core issue can be stated as follows: in theories where the Hamiltonian that generates the evolution of a system is a constraint, the natural interpretation of the theory is that there is no physical change with respect to time. Unpacking this problem highlights several subtleties concerning the construction and interpretation of the constrained Hamiltonian formalism, and therefore provides a helpful starting point for setting up the theses of this dissertation.

The history can be traced back to Dirac (1964), who presents the first account of the constrained Hamiltonian formalism. A crucial part of Dirac's account is the definition of a "gauge transformation" as a transformation generated by arbitrary combinations of the first-class constraints, where the first-class constraints are those constraints that have vanishing Poisson bracket with any constraint. The presence of gauge transformations indicates that there is non-uniqueness in the solutions to the equations of motion, in that there are multiple mathematically distinct evolutions from an initial state. In other words, the presence

[1] For some examples, see Earman (2002a); Maudlin (2002); Thébault (2012); Thébault (2021); Belot (2007); Gryb and Thébault (2023, 2016); Gryb and Thébault (2016a); Pitts (2014a).

of gauge transformations indicates that there is *indeterminism* in the theory. However, this indeterminism is regarded as only apparent; gauge transformations are interpreted as symmetries that connect physically equivalent descriptions of the same state or history of a system.

In theories where the Hamiltonian is itself a first-class constraint, this definition and interpretation of gauge transformations leads to the conclusion that the states along the solutions to the equations of motion of a theory are precisely the states that are physically equivalent. This is one of the expressions of the Problem of Time, which Thébault (2012) has dubbed the *problem of representing change*. Around the same time as Dirac, Bergmann (1961) provided the definition of an 'observable' as a phase space function that has a weakly vanishing Poisson bracket with the first-class constraints, and thus along with Dirac's notion of a gauge transformation, the observables are those functions that are gauge-invariant. A consequence of Dirac's definition of a gauge transformation in theories where the Hamiltonian is a first-class constraint is that the observables do not change over time. Therefore, it seems that the quantities that are naturally regarded as the physical ones, since they are gauge-invariant, do not have any dynamical evolution. This is another expression of the Problem of Time; the *problem of representing observables* (again, following Thébault (2012)).

The Problem of Time becomes even more prominent when translated into a geometric formulation of the constrained Hamiltonian formalism. In the geometric formulation, there is a natural way to 'remove' the gauge degrees of freedom by quotienting the state space by the points connected by gauge transformations (the 'gauge orbits') to give what is known as the 'reduced phase space'. On the reduced phase space, only the observables are definable as functions on this space. In the case of a Hamiltonian constraint, this process leads to a reduced phase space with trivial dynamics in the sense that we cannot describe the evolution of a system since the points of the reduced phase space are entire solutions.

There are several reasons why the Problem of Time is seen as a significant challenge for modern physics. First, a theory for which the Hamiltonian is a first-class constraint is the canonical formulation of General Relativity, the most successful theory of space and time. Second, the constrained Hamiltonian formalism is the basis for a standard technique of quantizing a classical theory, known as Dirac or canonical quantization. Using this technique to quantize General Relativity leads to the Wheeler-DeWitt equation "$H\psi = 0$", which is naturally interpreted as saying that the wavefunction of the universe, which encodes the physical degrees of freedom, has no evolution. In other words, the Problem of Time manifests itself when attempting to formulate a quantum theory of gravity, one of the central programs in modern physics. Thus, it appears that our best theories of physics provide a picture that is incompatible with our experience of time and change.

The Problem of Time is also connected to philosophical debates concerning when a theory should be characterized as having redundancy or "surplus structure", and how one ought to respond to the presence of surplus structure. One common view is that surplus structure exists when there is a particular kind of symmetry between models of a theory, and that removing surplus structure requires equivocating between symmetry-related models. Inasmuch as gauge transformations are a candidate symmetry of this kind, the Problem of Time can be seen as situated in this wider debate about how to characterize the symmetries that indicate surplus structure and what it would mean to equivocate between symmetry-related models in a way that preserves the empirical content of the theory.

There have been several 'solutions' to the Problem of Time in the literature. One response in the physics literature is to accept the timeless formalism but provide a picture where time 'emerges' from the fundamentally timeless picture.[2] Another response is to argue that the Dirac definition of gauge transformations does not apply to the Hamiltonian first-class constraint, since the Hamiltonian constraint has a distinct formal and interpretational role

[2] For an overview of this response, see Anderson (2012).

in comparison to other constraints.[3] Both responses agree that the standard definition and interpretation of gauge transformations in the constrained Hamiltonian formalism is unproblematic in general but argue that there is something special about the case of time; either we should see it as an emergent feature of our theory or prevent it from being lost in the first place.

In this dissertation, I approach the problem from another angle: I ask whether the standard way of formulating and interpreting the constrained Hamiltonian formalism is correct. Indeed, there have been some recent challenges put forward against the orthodox view. For one, some authors have argued that it is not the case that *arbitrary* combinations of first-class constraints generate gauge transformations, against the Dirac definition.[4] Second, the claim that gauge variables do not correspond to anything 'physical' has been disputed; a prominent response of this kind is the partial observables approach pioneered by Carlo Rovelli, which provides a picture where the gauge variables have an interpretation as the measurable quantities.[5] Therefore, there are reasons to think that the orthodoxy is in trouble even outside of cases that fall under the Problem of Time, and it is these troubles with which we concern ourselves here.

In more detail, this dissertation asks the following questions: (1) What is the correct characterization of the gauge transformations in the constrained Hamiltonian formalism? (2) Does this characterization 'match' the characterization given in the Lagrangian formalism? (3) Should we regard gauge transformations as indicating 'redundancy' in the mathematical formalism? A theme running through the dissertation is that progress can be made on these questions by making precise the geometric structure of a constrained Hamiltonian theory. I

[3]For responses of this kind, see Kuchař (1991), Barbour (1994), Barbour and Foster (2008), Gryb and Thébault (2016b).

[4]See, for example, Pons (2005); Pitts (2014a,b). Henneaux and Teitelboim (1994) §1.2.2, §1.6.3 also present supposed counterexamples to the claim that secondary first-class constraints generate gauge transformations.

[5]For an overview of this approach, see Rovelli (2004, 2002, 2014). For more details, see Dittrich (2006, 2007).

argue that this precise characterization can (1) be used to resolve a debate about how to correctly characterize the relationship between constraints and gauge transformations, (2) shed light on the relationship between the constrained Hamiltonian formalism and the Lagrangian formalism, and (3) provide an avenue for formulating a constrained Hamiltonian theory that resolves an apparent tension between gauge variables being 'excess structure' and playing an ineliminable role.

This dissertation will therefore not directly answer the question of how to resolve the Problem of Time. However, I think that exploring the foundations of the constrained Hamiltonian formalism in the 'orthodox' cases/cases where the Problem of Time does not arise, provides progress on the question of how the Problem of Time should be characterized and understood. In particular, the picture provided in this dissertation will suggest the following. First, one part of the special status of the Hamiltonian first-class constraint is that one cannot clearly distinguish the notion of a gauge transformation on *states* and the notion of a gauge transformation on *solutions* generated by it.[6] Second, the Problem of Time is not the only case where moving to the reduced phase space comes with interpretational issues. Finally, inasmuch as the Problem of Time is a problem in the constrained Hamiltonian formalism, it is equally a problem in the corresponding Lagrangian formalism. These suggest that considering the Problem of Time in the broader framework of how to correctly characterize the constrained Hamiltonian formalism, its symmetries, and its relationship to the Lagrangian formalism provides new perspectives on the issue that are worth exploring further.

There are effectively two parts to the dissertation. The first part (Chapters 1 and 2) considers the *definition* of the gauge transformations. The second part (Chapters 3 and 4) considers the *interpretation* of gauge transformations. These two parts are intertwined with one another; my argument for what the correct definition of the gauge transformations is will depend

[6] The bearing of the distinction between gauge transformations on states and gauge transformations on solutions for the Problem of Time is discussed further in Gryb and Thébault (2023).

partly on my views about how to interpret physical theories more generally. Similarly, my argument for how to interpret the gauge transformations will make use of my views about the correct definition of the gauge transformations. However, I take these projects to be distinct: one could be convinced of my argument for what the correct definition of the gauge transformations is without being convinced of my argument for the interpretation of gauge transformations, and vice versa.

In Chapter 1, I consider whether the standard definition of a gauge transformation in the constrained Hamiltonian formalism is correct. On the basis of the definition that a gauge transformation is a transformation generated by an arbitrary combination of first-class constraints, Dirac argued that one should extend the form of the Hamiltonian in order to include all of the gauge freedom. However, there have been some recent dissenters of Dirac's view. Notably, Pitts (2014b) argues that a first-class constraint can generate "a bad physical change" and therefore that extending the Hamiltonian in the way suggested by Dirac is unmotivated. In this chapter, I use a geometric formulation of the constrained Hamiltonian formalism to argue that there is a flaw in the reasoning used by both sides of the debate, but that correct reasoning supports the standard definition and the extension to the Hamiltonian. In doing so, I clarify two conceptually different ways of understanding gauge transformations, and I pinpoint what it would take to deny that the standard definition is correct.

In Chapter 2, I consider a possible counterargument to the argument of Chapter 1: that the standard definition in the constrained Hamiltonian formalism is at odds with the definition provided by the Lagrangian formalism, and therefore that this renders them inequivalent theories. I argue that this argument relies on a particular formulation of the Lagrangian formalism, and that just as one can motivate the extension to the form of the Hamiltonian in the context of the geometric formulation of Hamiltonian mechanics, one can motivate a similar extension to the Lagrangian in the geometric formulation of the Lagrangian formalism. I show that this reformulation of the Lagrangian formalism is equivalent to the extended Hamiltonian formalism under a particular characterization of the structure of these theories.

In Chapter 3, I turn to the interpretational question of when one should regard a theory as having "excess structure". I present a distinction between two kinds of structure that I call *theoretical structure* and *auxiliary structure*, and I argue that understanding the distinctive role that each structure plays helps to discriminate between different ways of removing excess structure.

In Chapter 4, I consider how the argument in Chapter 3 bears on the constrained Hamiltonian formalism. I argue that the literature regarding the interpretation of gauge transformations conflates the distinction between theoretical structure and auxiliary structure, and that this leads to mistakenly thinking that the only option for removing excess structure in the constrained Hamiltonian formalism is moving to the reduced phase space. Moreover, I argue that one can reconstruct the partial observables approach to gauge variables as providing an argument in favor of an alternative way of removing excess structure in the constrained Hamiltonian formalism.

Together, these chapters can be understood as showing that part of formulating precisely the mathematical structure of a physical theory is an interpretational task. This is not a new idea; indeed, much of this dissertation is inspired by work on the use of formal tools to spell out notions such as 'theoretical equivalence' and 'excess structure'.[7] However, what sets this dissertation apart is that (1) these tools have not been put to work explicitly in the context of the constrained Hamiltonian formalism and (2) I argue that there is an underappreciated way in which interpretation plays a role in determining how to remove excess structure from a theory that the constrained Hamiltonian formalism and its associated puzzles helps to draw out.

[7]A (not comprehensive) list of some of the work on the use of formal tools to spell out notions of equivalence and excess structure that has inspired this dissertation are Barrett (2015a,b, 2019); Weatherall (2016a,b, 2018, 2019a,b); Halvorson (2012, 2016); Dewar (2019, 2022).

Chapter 1

Do First-Class Constraints Generate Gauge Transformations? A Geometric Resolution.

1.1 Introduction

Gauge transformations represent local symmetries in physics that are often taken to indicate arbitrariness in the mathematical formalism of a theory. How to interpret this arbitrariness is widely disputed and is connected to a wider literature on "surplus structure" in physics.[1] However, there is a different kind of dispute about gauge transformations that will be the focus here: if gauge transformations are conceptualized as transformations that indicate arbitrariness, what is the correct *formal* definition of a gauge transformation?

[1]For more on the notion of surplus structure and its connection to symmetries of a theory, see, for example, Ismael and Van Fraassen (2003); Earman (2004); Baker (2010).

There is a longstanding tradition of using a formalism known as the “constrained Hamiltonian formalism" to establish the gauge transformations of a theory. The standard definition arising from this formalism is attributed to Dirac (1964): a gauge transformation is a transformation generated by an arbitrary combination of first-class constraints, which are the constraints on the dynamically allowed states that have vanishing Poisson bracket with all of the constraints. This definition is taken to have important consequences for the formulation of a Hamiltonian theory. In particular, Dirac argued on the basis of this definition that the Hamiltonian function that generates the dynamics should be understood as an equivalence class of Hamiltonians called the “Extended Hamiltonian".

However, there have been several recent dissenters of Dirac's account of gauge transformations. For example, using the case of Electromagnetism, Pitts (2014b) argues that a first-class constraint can generate “a bad physical change". Similarly, Pons (2005) argues that Dirac's analysis of gauge transformations is “incomplete" since it does not provide an accurate account of the symmetries between solutions to the equations of motion. Both authors conclude that formulating a theory in terms of the Extended Hamiltonian is unmotivated. If correct, these arguments could have implications for other issues in the foundations of the constrained Hamiltonian formalism. Notably, there is a puzzle called the “Problem of Time" that arises in the constrained Hamiltonian formalism for theories that are time-reparameterization invariant when one adopts the standard definition of a gauge transformation. If gauge transformations are not given by the standard definition, then this could be an avenue to avoiding the Problem of Time.[2]

More recently, Pooley and Wallace (2022) argue, contra Pitts (2014b), that one can vindicate Dirac's orthodoxy in the case of Electromagnetism by showing that if one formulates the theory by starting with the Extended Hamiltonian, then this formulation of the theory has the same empirical predictions as the alternative and arbitrary combinations of first-class con-

[2]See Pitts (2014a) for a response of this kind. For an introduction to the Problem of Time and its philosophical implications, see Thébault (2021).

straints generate gauge transformations. In this chapter, I extend the observations of Pooley and Wallace (2022) by arguing the stronger and more general claim that the formulation of a gauge theory that treats the Extended Hamiltonian as the equivalence class of Hamiltonians is motivated on theoretical grounds. In more detail, I use a standard geometric formulation of the constrained Hamiltonian formalism to show that the Extended Hamiltonian can be motivated independently from consideration of the gauge transformations, and, under the dynamics generated by the Extended Hamiltonian, the standard account of gauge transformations as being generated by arbitrary combinations of first-class constraints is correct. In doing so, I argue that there is a common assumption made in the literature about the relationship between gauge transformations and the form of the Hamiltonian that is unnatural in the geometric framework. This leads to a revised account of the definition of gauge transformations in the constrained Hamiltonian formalism that sheds light on a particular source of contention: what the relationship is between gauge transformations on *states* and gauge transformations on *solutions*.

The chapter will go as follows. In Section 1.2, I present Dirac's version of the constrained Hamiltonian formalism and his argument that arbitrary combinations of first-class constraint generate gauge transformations. In Section 1.3, I spell out the example that Pitts (2014b) gives as a counterexample to Dirac's view. In Section 1.4, I discuss where the disagreement lies between Dirac and Pitts' views, and I highlight a crucial assumption made on both sides of the debate. In Section 1.5, I consider the response to Pitts (2014b) given by Pooley and Wallace (2022) and pinpoint the way in which it fails to provide a complete response. In Section 1.6, I present the geometric formulation of the constrained Hamiltonian formalism, and I use this formulation in Section 1.7 to argue that the issue in the debate lies in the way that gauge transformations are both understood and motivated. In Section 1.8, I consider two possible counterarguments, before concluding.

1.2 Dirac's Theory

Dirac's version of the constrained Hamiltonian formalism is constructed by starting with the Lagrangian formalism. In the Lagrangian framework, one has a finite N number of degrees of freedom $q_n, n = 1, ..., N$, with corresponding velocities $\frac{dq_n}{dt} = \dot{q}_n$, where we assume an independent time variable t.[3] The dynamics are given by specifying a Lagrangian $L = L(q_n, \dot{q}_n)$ with corresponding action $I = \int L(q_n, \dot{q}_n)dt$, from which one derives the equations of motion called the Euler-Lagrange equations:

$$\frac{d}{dt}\frac{\partial L(q_n, \dot{q}_n)}{\partial \dot{q}_n} = \frac{\partial L(q_n, \dot{q}_n)}{\partial q_n}$$

To move to the Hamiltonian framework, one introduces "canonical momenta" variables $p_n = \frac{\partial L}{\partial \dot{q}_n}$. When these momenta are not independent of each other, there are constraints of the form $\phi_m(q_n, p_n) \approx 0$ for $m = 1, ..., M$ where M is the number of constraints and the equality is *weak* equality, indicating that the constraints only hold on a subspace of phase space (the state space given by the collection of points (q_n, p_n)). Constraints of this kind are called the *primary constraints.*

The *Hamiltonian* is defined as $H(q_n, p_n) = p^n q_n - L$ where the upper and lower indices indicates a sum over n. However, it is not uniquely defined when the system is constrained, since one can add a linear combination of primary constraints and it will weakly be the same Hamiltonian. We call the addition of this linear combination of primary constraints the *Total Hamiltonian*, $H_T = H + u^m \phi_m$ where u^m are arbitrary functions of the canonical variables and again we implicitly have a sum over m. The Total Hamiltonian should therefore be thought of as an equivalence class of Hamiltonians, differing over the choices of u^m. From the variation in H_T, one can derive Hamilton's equations of motion with constraints:

[3] In order to consider the Problem of Time, it is useful to drop this assumption and treat the time variable as an additional dynamical variable, but we keep this assumption for the purposes here.

$$\dot{q}_n = \frac{\partial H}{\partial p_n} + u^m \frac{\partial \phi_m}{\partial p_n}$$

$$\dot{p}_n = -\frac{\partial H}{\partial q_n} - u^m \frac{\partial \phi_m}{\partial q_n}$$

More generally, for any dynamical variable g, $\dot{g} \approx \{g, H\} + u^m\{g, \phi_m\} = \{g, H_T\}$ where $\{\}$ is the Poisson bracket.[4]

In order for the solutions to the equations of motion to be consistent with the primary constraints, in the sense that the primary constraints hold at all times along a solution to the equations of motion, it ought to be the case that $\dot{\phi}_m \approx 0$. In other words, it ought to be the case that $\{\phi_m, H\} + u^{m'}\{\phi_m, \phi_{m'}\} \approx 0$. For each m, this equation either is identically satisfied with the primary constraints, reduces to an equation independent of the u's of the form $\chi_k(q_n, p_n) \approx 0$, or it imposes conditions on the u's.

In the second case, we say that $\chi_k(q_n, p_n) \approx 0$ are *secondary constraints*, since they arise from applying the equations of motion to the primary constraints. If we have a secondary constraint, then we get new consistency conditions by requiring $\dot{\chi}_k \approx 0$, which is again one of the three kinds above. One can continue this process until one has found all of the secondary constraints and one is left with the consistency conditions of the third kind. We can combine the primary and secondary constraints, writing them as $\phi_j \approx 0$ for $j = 1, ..., M + K$ where K is the number of secondary constraints.

For the remaining consistency conditions that do not reduce, we find solutions $u^m = U^m + v^a V_a^m$ where v^a is arbitrary and $V^m\{\phi_j, \phi_m\} \approx 0$. Substituting into the Total Hamiltonian, we get

[4]The Poisson bracket satisfies the following properties: 1. If k is a constant, then for any function f, $\{k, f\} = 0$. 2. Leibniz rule: for any functions f, g, h, $\{fg, h\} = f\{g, h\} + g\{f, h\}$. 3. Jacobi identity: for any functions f, g, h, $\{f, \{g, h\}\} + \{g, \{h, f\}\} + \{h, \{f, g\}\} = 0$.

$$H_T = H' + v^a \phi_a$$

where $H' = H + U^m \phi_m$ and $\phi_a = V_a^m \phi_m$. Notice that we have satisfied all the consistency conditions but still have coefficients v^a that are arbitrary functions of the canonical variables.

A dynamical variable $R(q_n, p_n)$ is said to be *first-class* if $\{R, \phi_j\} \approx 0$. In other words, a dynamical variable is first-class if the Poisson bracket with any constraint equals a linear function of the constraints. If it is not first-class, it is called *second-class*. Importantly, H' and ϕ_a are first-class. This means that H_T is an equivalence class of Hamiltonians given by a sum of a first-class Hamiltonian and a linear combination of primary, first-class constraints.

Given some initial state $(q_n(t_0), p_n(t_0))$, the q's and p's at later times are underdetermined because of the arbitrariness in the coefficients v^a. One might take this to be a mark of *indeterminism* in the theory: there are multiple possible evolutions from an initial state. However, we might also think that this indeterminism is an artifact of our mathematical description, in that it indicates that our theory contains "redundancy". It is this direction of thought that led Dirac to propose the following definition of a gauge transformation:

> **State Gauge Transformation**: A gauge transformation relates any two states that are possible evolutions from an initial state under the dynamics generated by the Total Hamiltonian at some fixed (infinitesimal) interval δt.

In other words, Dirac proposes that physically equivalent states as precisely those that result from the arbitrariness in v^a in evolving the state of a system.

We can determine these transformations in the following way. For a given dynamical variable g with initial value g_0, its value after some infinitesimal δt under a specified choice of coefficients v^a is:

$$g(\delta t) = g_0 + \dot{g}\delta t = g_0 + \{g, H_T\}\delta t = g_0 + \delta t[\{g, H'\} + v^a\{g, \phi_a\}] \tag{1.1}$$

However, one could have made different choices for v^a. Call another set of choices v'^a. The difference between the two values for g at δt under these two choices of coefficients is given by:

$$\Delta g(\delta t) = \delta t(v^a - v'^a)\{g, \phi_a\} = \varepsilon^a\{g, \phi_a\} \tag{1.2}$$

where ε^a is an arbitrary small number. This change will describe the same physical state, since it corresponds to a change from one state to another that arises merely from a different choice of arbitrary coefficient in the evolution from some initial state. Since ϕ_a are just the primary first-class constraints, Dirac concludes:

> All primary first-class constraints generate gauge transformations.

However, this isn't the end of the story. Take some value for $g(\delta t)$ and transform it by $\varepsilon^a\{g, \phi_a\}$ *twice*. This new value for $g(\delta t)$ is related to the previous value by some amount generated by $\{\phi_a, \phi_{a'}\}$. The ϕ_a's are first-class constraints, and the Poisson bracket of two first-class quantities is first-class, so this generating function is a first-class constraint. However, it need not be a primary first-class constraint; it could be a *secondary* first-class constraint. Observing this, Dirac presents the following conjecture:

> **Dirac Conjecture**: All secondary first-class constraints generate gauge transformations.

We therefore conclude that:

> Arbitrary combinations of first-class constraints generate a State Gauge Transformation.

However, we are now in a situation where the dynamics are given by the Total Hamiltonian, which includes the arbitrariness associated with the primary first-class constraints. On the other hand, we also have arbitrariness associated with the secondary first-class constraints. This mismatch between the dynamics and the arbitrariness led Dirac to suggest that one should also add the first-class *secondary* constraints to the Total Hamiltonian, giving rise to the *Extended Hamiltonian*, $H_E = H_T + w^b \chi_b$ where χ_b are the first-class secondary constraints and w^b are arbitrary functions of the canonical variables. The equations of motion then read: $\dot{g} = \{g, H_E\}$.

Finally, we define an *observable* as a function f that has the property that $\{f, \varphi_j\} \approx 0$ for all first-class constraints φ_j. Observables are functions that are gauge-invariant, in the sense that they take the same value under the transformations generated by the first-class constraints. On the other hand, the *gauge variables* are the functions that are not observables.

The final picture of Dirac's theory is:

1. The symmetries of the theory are “State Gauge Transformations" that are generated by arbitrary combinations of first-class constraints.

2. The dynamics are generated by an equivalence class of Hamiltonians represented by the Extended Hamiltonian.

Whether this picture is correct will be the subject of the rest of the chapter.

1.3 An Argument Against Dirac

Although Dirac's account of the gauge transformations in the constrained Hamiltonian formalism has been widely accepted as the standard framework, there are recent arguments that Dirac's account is flawed.[5] Here, I focus on the argument by Pitts (2014b) that classical Electromagnetism provides a counterexample to Dirac's account.

The Lagrangian for classical Electromagnetism can be written in observer-dependent form as

$$\mathcal{L}(\vec{A}, V; \dot{\vec{A}}, \dot{V}) = \int \frac{1}{2}(\dot{\vec{A}} - \nabla V)^2 - \frac{1}{2}(\nabla \times \vec{A})^2 - (V\rho + \vec{A} \cdot \vec{J})$$

where $\vec{A}$ and V are time-dependent functions on $\mathbb{R}^3$ and the integral is over $\mathbb{R}^3$. The conjugate momenta are $p_{\vec{A}} = \frac{\delta L}{\delta \dot{\vec{A}}} = \dot{\vec{A}} - \nabla V$ and $p_V = \frac{\delta L}{\delta \dot{V}} = 0$. This means that there is one primary constraint, $\phi_0 = p_V$. The Total Hamiltonian is:

$$H_T = \int \frac{1}{2}(p_{\vec{A}}^2 + \vec{B}^2) + \lambda p_V + p_{\vec{A}} \cdot \nabla V + (V\rho + \vec{A} \cdot \vec{J}) \tag{1.3}$$

where the integral is over $\mathbb{R}^3$ and λ is an arbitrary function of the canonical coordinates. Integrating by parts with appropriate boundary conditions, we can rewrite the Total Hamiltonian as:

$$H_T = \int \frac{1}{2}(p_{\vec{A}}^2 + \vec{B}^2) + \vec{A} \cdot \vec{J} + \lambda p_V - V(\nabla \cdot p_{\vec{A}} - \rho) \tag{1.4}$$

[5] See in particular Pitts (2014a,b) and Pons (2005) but also Pons et al. (1997) and Barbour and Foster (2008).

We can then find the evolution of the primary constraint:

$$\{p_V, H_T\} = \frac{\delta H}{\delta V} = \nabla \cdot p_{\vec{A}} - \rho. \tag{1.5}$$

So there is a secondary constraint given by $\phi_1 = \nabla \cdot p_{\vec{A}} - \rho$. The evolution of the secondary constraint is zero, so there are two constraints in total, and both constraints are first-class.

The equations of motion for $\vec{A}$ and V are given by:[6]

$$\begin{aligned} \frac{\partial \vec{A}}{\partial t} &= \{\vec{A}, H_T\} = \frac{\partial H_T}{\partial p_{\vec{A}}} = p_{\vec{A}} + \nabla V \\ \frac{\partial V}{\partial t} &= \{V, H_T\} = \frac{\partial H_T}{\partial p_V} = \lambda \end{aligned} \tag{1.6}$$

The question that Pitts (2014b) asks is whether the arbitrary combinations of the primary and secondary constraint generate gauge transformations for these equations. In other words, we want to know whether, if $(\vec{A}(t), V(t); p_{\vec{A}}(t), p_V(t))$ satisfies these equations of motion, then transforming this solution by an arbitrary combination of the first-class constraints, $\int \alpha\phi_0 + \beta\phi_1$, also satisfies the equations of motion, where α and β are arbitrary functions of the canonical coordinates and time.

We have that:

[6] We leave out the equations of motion for $p_{\vec{A}}$ and p_V for convenience, since they aren't important for the argument.

$$
\begin{aligned}
\{\vec{A}, \int \alpha\phi_0 + \beta\phi_1\} &= \{\vec{A}, \int \alpha p_V + \beta(\nabla \cdot p_{\vec{A}} - \rho)\} \\
&= \{\vec{A}, \int \alpha p_V\} + \{\vec{A}, \int \beta(\nabla \cdot p_{\vec{A}} - \rho)\}
\end{aligned}
\tag{1.7}
$$

The first term vanishes. Since $\int \beta\nabla \cdot p_{\vec{A}} = -\int p_{\vec{A}} \cdot \nabla\beta$ by integration by parts (with appropriate boundary conditions), the second term is equal to $\{\vec{A}, -\int p_{\vec{A}} \cdot \nabla\beta + \beta\rho)\} = \nabla\beta$. Therefore, the transformed quantity is given by $A' = A + \nabla\beta$.

Similarly:

$$
\{V, \int \alpha\phi_0 + \beta\phi_1\} = \{V, \int \alpha p_V\} + \{V, \int \beta(\nabla \cdot p_{\vec{A}} - \rho)\} \tag{1.8}
$$

The second term here vanishes, and the first term is equal to α. Thus, the transformed potential is given by $V' = V + \alpha$.

We also have that $\{p_{\vec{A}}, \int \alpha p_V + \beta(\nabla \cdot p_{\vec{A}} - \rho)\} = \{p_V, \int \alpha p_V + \beta(\nabla \cdot p_{\vec{A}} - \rho)\} = 0$ and so the conjugate momenta do not change under the transformation generated by an arbitrary combination of the constraints. We can therefore write the transformed equations of motion for $\vec{A}$ and V as:

$$
\begin{aligned}
&\frac{\partial \vec{A}'}{\partial t} = \frac{\partial \vec{A}}{\partial t} + \frac{\partial \nabla\beta}{\partial t} = p_{\vec{A}} + \nabla(V + \alpha) \\
&\frac{\partial V'}{\partial t} = \frac{\partial V}{\partial t} + \frac{\partial \alpha}{\partial t} = \lambda
\end{aligned}
\tag{1.9}
$$

Since we assumed that $\frac{\partial \vec{A}}{\partial t} = p_{\vec{A}} + \nabla V$, the first equation is satisfied only when $\frac{\partial \nabla \beta}{\partial t} - \nabla \alpha = 0$. In particular, in the case where either α or β is zero (where one considers the transformation generated by only one of the primary or secondary constraints), the first equation is not satisfied.

On the basis of this argument, Pitts (2014b) concludes that arbitrary combinations of first-class constraints do not generate gauge transformations. Rather, only a particular combination of first-class constraints generates a gauge transformation. So, the argument goes, Dirac was wrong about what the gauge transformations are.

Remember also that the motivation for Dirac to move to the Extended Hamiltonian was that all secondary first-class constraints generate gauge transformations in addition to primary first-class constraints. But the above argument suggests that this is not true. In fact, it suggests there are only as many arbitrary functions of time as there are primary first-class constraints. To see this, notice that since $\nabla \alpha = \frac{\partial \nabla \beta}{\partial t}$, we can write the gauge transformations as being generated by $\int \dot{\epsilon}\phi_0 + \epsilon\phi_1$. In other words, we only need one arbitrary function (and its time derivative) to specify the gauge transformations. Therefore, one might also take this argument to show that the Extended Hamiltonian is not motivated. More strongly, it suggests that the Extended Hamiltonian is the *wrong* equivalence class of Hamiltonians since the Extended Hamiltonian gives rise to "more" arbitrariness in the dynamics than there in fact is, while the Total Hamiltonian captures exactly the arbitrariness in the dynamics.

1.4 Where The Disagreement Lies

There is an immediate sense in which the above argument fails on its own to show that Dirac was wrong. In Section 2, we interpreted Dirac as giving an account of what I called "State Gauge Transformation": transformations relating two states that are possible evolutions from some initial state. However, the argument I just ran, following Pitts (2014b), doesn't consider whether two *states* are equivalent; it considers whether two *solutions* are equivalent. That is, it considers whether arbitrary combinations of first-class constraints generate a transformation that takes one from a solution to the equations of motion to another solution. We might alternatively call this notion of a gauge transformation "Solution Gauge Transformation":

> **Solution Gauge Transformation**: A gauge transformation relates any two *curves* that are possible evolutions from an initial state under the dynamics generated by the Total Hamiltonian.

What Pitts' argument demonstrates is that the Solution Gauge Transformations are not generated by arbitrary combinations of first-class constraints in the context of classical Electromagnetism. Indeed, arbitrary combinations of first-class constraints *do* generate State Gauge Transformations in classical Electromagnetism. To see this, recall that we can write the Solution Gauge Transformations as $\int \dot{\epsilon}\phi_0 + \epsilon\phi_1$. At a particular fixed time, ϵ and $\dot{\epsilon}$ become independent of each other. And so, we can write the State Gauge Transformations as $\int \alpha\phi_0 + \beta\phi_1$, as would be the case if arbitrary combinations of first-class constraints generate gauge transformations. So what Pitts (2014b) shows is that Solution Gauge Transformations do not always match the State Gauge Transformations.

At this point one might want to say: what this shows is that we really have two distinct notions of a gauge transformation, 'State Gauge Transformation' and 'Solution Gauge Transformation', and it turns out that these notions do not coincide. This would suggest that there is not really a debate here at all; different parties in the debate are just focusing on different notions, and we can accept that both are right.

Although *formally* this thought seems correct, there is a *conceptual* issue with accepting both notions of a gauge transformation, since it would mean accepting that individual states along two curves can be gauge-equivalent without it being the case that if one curve is a solution, then the other also is. The reason is that the transformations that generate Solution Gauge Transformations are more restrictive than (are a subset of) those that generate State Gauge Transformations. But if gauge equivalence is supposed to mean *physical* equivalence, then this would be to say that two curves can be such that each individual state along one curve is physically equivalent to a state along the other curve but the curves *as a whole* are not physically equivalent to one another. Conceptually, this is not coherent: solutions just consist of a series of states, and so if all of these states are physically equivalent to some other series of states, then the solutions ought to also be physically equivalent.

Therefore, it seems that if one wants to accept that "Solution Gauge Transformation" is the right definition of gauge transformations on solutions and that gauge equivalence is a notion of physical equivalence, one has to accept that there is no independent notion of a gauge transformation on states. That is, any notion of a state gauge transformation must be derivative to that of the solution gauge transformations: a state gauge transformation must be the special case of the solution gauge transformations where the solutions are considered to be infinitesimally short in terms of time.

This helps to set up the rest of the chapter: I will argue that one can maintain separate notions of state and solution gauge transformations as notions of physical equivalence, but it means that one has to deny that "State Gauge Transformation" and "Solution Gauge Transformation" as I defined them above are the right characterizations of gauge transformations on states and solutions respectively. In particular, one common part of the definition "State Gauge Transformation" and "Solution Gauge Transformation" is the commitment to gauge transformations being determined by considering curves that are generated by the *Total Hamiltonian.* I will argue that 1. gauge transformations on states do not require a commitment to a particular form of the Hamiltonian and 2. gauge transformations on solutions ought to be determined by considering curves that are generated by the Extended Hamiltonian rather than the Total Hamiltonian. This second argument bears a close resemblance to a recent response to Pitts (2014b) by Pooley and Wallace (2022), so it will be helpful to spell out their argument first and pinpoint the way in which it falls short of providing a complete resolution to the debate before detailing the two arguments.

1.5 A Response to Pitts

Pooley and Wallace (2022) show that in the example of classical Electromagnetism, if one starts with the Extended Hamiltonian, arbitrary combinations of first class constraints generate gauge transformations of solutions. Their argument can be summarised as follows. Consider the Extended Hamiltonian for classical Electromagnetism, where we add to the Total Hamiltonian the secondary constraint multiplied by an arbitrary function μ:

$$H_E = \int \frac{1}{2}(p_{\vec{A}}^2 + \vec{B}^2) + \vec{A}\cdot\vec{J} + \lambda p_V - (V+\mu)(\nabla\cdot p_{\vec{A}} - \rho) \tag{1.10}$$

With this Hamiltonian, the equations of motion become:

$$\begin{aligned} \frac{\partial\vec{A}}{\partial t} &= \frac{\partial H_E}{\partial p_{\vec{A}}} = p_{\vec{A}} + \nabla(V+\mu) \\ \frac{\partial V}{\partial t} &= \frac{\partial H_E}{\partial p_V} = \lambda \end{aligned} \tag{1.11}$$

When we now consider the transformation generated by an arbitrary combination of primary and secondary constraints, $\int \alpha\phi_0 + \beta\phi_1$, we find:

$$\begin{aligned}\frac{\partial \vec{A}'}{\partial t} &= \frac{\partial \vec{A}}{\partial t} + \frac{\partial \nabla \beta}{\partial t} = p_{\vec{A}} + \nabla(V + \mu + \alpha) \\ \frac{\partial V'}{\partial t} &= \frac{\partial V}{\partial t} + \frac{\partial \alpha}{\partial t} = \lambda\end{aligned} \tag{1.12}$$

We can rewrite the first equation as $\frac{\partial \vec{A}'}{\partial t} = \frac{\partial \vec{A}}{\partial t} = p_{\vec{A}} + \nabla(V + \mu + \alpha - \dot{\beta})$. Notice that μ, α and $\dot{\beta}$ are all arbitrary functions, so we can write this equation as

$$\frac{\partial \vec{A}'}{\partial t} = \frac{\partial \vec{A}}{\partial t} = p_{\vec{A}} + \nabla(V + \mu')$$

where μ' is arbitrary. This is just the untransformed equation of motion, with μ' in place of μ. In other words, if $(\vec{A}(t), V(t); p_{\vec{A}}(t), p_V(t))$ is a solution to $\frac{\partial \vec{A}}{\partial t} = p_{\vec{A}} + \nabla(V + \mu)$, then $(\vec{A}(t) + \nabla\beta, V(t) + \alpha; p_{\vec{A}}(t), p_V(t))$ is also a solution. Therefore, arbitrary combinations of first-class constraints generate gauge transformations on solutions, for the dynamics generated by the Extended Hamiltonian.

Although this argument shows that when we start with the Extended Hamiltonian, the gauge transformations are generated by arbitrary combinations of first-class constraints, it leaves open the question of what the justification is for starting with the Extended Hamiltonian. Indeed, it seems that the proponents of "Solution Gauge Transformation" will deny that this is the right starting point; they would say that it is the Total Hamiltonian that one should use to determine the gauge transformations.

Pooley and Wallace (2022) do provide one kind of response: the dynamics generated by the Extended Hamiltonian is *empirically equivalent* to the dynamics generated by the Total Hamiltonian, in the sense that the predictions regarding gauge-invariant quantities (the observables) are the same. In particular, what they notice is that the difference between the solutions of the Total and Extended Hamiltonian lies in what quantity plays the role of the electric field: when the Total Hamiltonian is used to generate the dynamics, it is $\dot{\vec{A}} - \nabla V$ that plays the role of the electric field, but when the Extended Hamiltonian is used, it is $p_{\vec{A}}$. And so, given that our access to these quantities is through the role they play in the equations of motion, there is no empirical difference between these choices of Hamiltonian.

Although I take this response to be both convincing and informative, I will argue that we can go further: the Extended Hamiltonian can be motivated purely on mathematical grounds, and therefore there are *theoretical* reasons for using the Extended Hamiltonian to determine the gauge transformations.

To make this argument, I will use a standard geometric way of expressing the constrained Hamiltonian formalism since it provides a neutral framework for illuminating the issues of concern. In particular, the geometric framework allows us to see clearly what the role of the first-class constraints is within the structure of the formalism. This will help to make clear the sense in which there are theoretical motivations for particular definitions of state and solution gauge transformations.

1.6 Geometric Formulation

The constrained Hamiltonian formalism can be expressed naturally in a geometric way using the theory of symplectic manifolds.[7] A symplectic manifold consists of a pair (M, ω) where M is a smooth manifold and ω is a *symplectic form*: it is a two-form (a smooth, anti-symmetric tensor field of rank (0,2)), that satisfies the following conditions:

1. ω is *non-degenerate*, i.e. if $\omega(X_i, X_j) = 0$ for all $X_j \in TM$ and some $X_i \in TM$, then $X_i = 0$.

[7] This formalism is widely used to express the constrained Hamiltonian formalism. For further details of this formalism, see Henneaux and Teitelboim (1994); Butterfield (2006).

2. ω is *closed*, i.e., $\mathbf{d}\omega = \mathbf{0}$, where $\mathbf{d}$ is the exterior derivative operator, which is such that $\mathbf{d}f = df$, the differential of a function f, $\mathbf{d}(\mathbf{d}\alpha) = 0$ where α is a k-form, and $\mathbf{d}(f\alpha) = df \wedge \alpha + f\mathbf{d}\alpha$.

There is a sense in which every symplectic manifold comes equipped with "Poisson structure": Let (M, ω) be a symplectic manifold and $C^\infty(M)$ the space of smooth maps on M. In addition, let ω' be the inverse of ω (a smooth, anti-symmetric tensor field of rank $(2, 0)$).[8] Then the map $\{\cdot, \cdot\} : C^\infty(M) \times C^\infty(M) \to C^\infty(M)$ defined by $f, g \mapsto \{f, g\} = \omega'(df)(dg)$ is a Poisson bracket on M.

A constrained Hamiltonian theory can be defined as a symplectic manifold in the following way. The manifold is the cotangent bundle of configuration space (otherwise known as phase space), T^*Q, whose points can be written as $\{(q_n, p_n), n = 1, ..., N\}$. T^*Q comes equipped with a one-form, the *Poincaré one-form*, given by $\theta = p_i dq^i$. The corresponding two-form is given by $\omega = \mathbf{d}\theta = dp_b \wedge dq^b$, which is symplectic.

Given a function f, one can uniquely define a smooth tangent vector field X_f through:

$$\omega(X_f, \cdot) = \mathbf{d}f \tag{1.13}$$

where $\{\cdot\}$ represents any vector field tangent to T^*Q. In particular, one can uniquely define a vector field corresponding to the Hamiltonian $H = p^i q_i - L$ through $\omega(X_H, \cdot) = dH$. This provides an alternative way to write Hamilton's equations. In particular, $\{f, H\} = \omega(X_f, X_H) = df(X_H) = \mathcal{L}_{X_H}(f)$. If we define the flow parameter of X_H to be time, then this says that $\{f, H\} = \frac{df}{dt}$, which is Hamilton's equation.

[8] This is well-defined because ω is non-degenerate.

We can understand the primary constraints $\phi_m(q_n, p_n) = 0$ for $j = 1, .., M$ where M is the total number of constraints as giving rise to a smooth, embedded sub-manifold of phase space of dimension $N - M$, which we call the *primary constraint surface*, given by $\Sigma_p = \{(q_n, p_n) \in \Gamma | \forall_m : \phi_m(q_n, p_n) = 0\}$. The first-class primary constraints are those constraints whose associated vector field is tangent to Σ_p, while the second-class primary constraints are those constraints whose associated vector field is not tangent to Σ_p. For the purposes here, we will restrict ourselves to the case where we just have first-class constraints, since these are the relevant ones for defining the gauge transformations.

We can define an induced two-form on the primary constraint surface $\tilde{\omega}_p$ as the pullback along the embedding $i : \Sigma_p \to \Gamma$ of ω. This induced two-form is in general *degenerate* i.e. it is not invertible. In particular, it possesses M linearly independent null vector fields that form the null space of $\tilde{\omega}_p$. These are the vector fields that satisfy $\tilde{\omega}(X_m, \cdot) = 0$ where $\{\cdot\}$ is any vector field tangent to Σ_p. But these are precisely the vector fields that off the constraint surface satisfy $\omega(X_m, \cdot) = d\phi_m$ where ϕ_m are the primary first-class constraints, since $d\phi_m|_{\Sigma_p} = 0$. Thus, we will write X_{ϕ_m} for these null vector fields to indicate that they are the tangent vector fields associated with the primary first-class constraints. The degeneracy of $\tilde{\omega}_p$ means that one cannot associate a unique vector field with any smooth function on the constraint surface through the equation $\tilde{\omega}_p(X_f, \cdot) = \mathbf{d}f$, since *if* X_f satisfies this equation (if it is tangent to the primary constraint surface), then so does $X_f + X_{\phi_m}$ since the two-form acts linearly.

We can write the equations of motion on the primary constraint surface as $\tilde{\omega}_p(X_H, \cdot) = dH|_{\Sigma_p}$. However, this equation of motion may not have solutions everywhere, since X_H may not be tangent to the primary constraint surface. In order for the solutions to be tangent to the primary constraint surface, it must be that $\tilde{\omega}_p(X_H, X_{\phi_m}) = dH(X_{\phi_m}) = 0$. This is geometrically what gives rise to the *secondary constraints*, and we can think of these secondary constraints as leading to the specification of a further submanifold.

Continuing this process of requiring the solutions to be tangent to the constraint surface terminates in some final constraint surface Σ_f, defined by the satisfaction of a collection of constraints $\varphi_j(q_n, p_n) = 0$ for $j = 1, ..., J$ where J is the total number of constraints. We can also define an induced two-form on Σ_f, $\tilde{\omega}_f$, whose null vector fields are the vector fields associated with all of the first-class constraints, which we will write as X_{φ_j} (since we are just considering the case where all the constraints are first-class, although it is easy to extend to the case where there are second-class constraints). The equations of motion are $\tilde{\omega}_f(X_H, \cdot) = dH|_{\Sigma_f}$, which has (non-unique) solutions everywhere on Σ_f.

The integral curves of the null vector fields are called the *gauge orbits*. Equivalently, the gauge orbits consist of the set of points that can be joined by a curve with null tangent vectors. The gauge orbits on the final constraint surface coincide with the notion of a gauge transformation in the Dirac formalism in the following sense: it is the null vector fields that generate the gauge orbits on the final constraint surface, and these coincide with the vector fields X_{φ_j} corresponding to the first-class constraints. And so, arbitrary combinations of first-class constraints effectively generate a transformation that takes one along the gauge orbits at each point.

We can also understand the *observables* in the geometric formulation as the functions that are constant along the gauge orbits. In other words, the observables are the functions f for which $\omega(X_f, X_{\varphi_j}) = 0$ on Σ_f, since $\omega(X_f, X_{\varphi_j}) = \mathcal{L}_{X_{\varphi_j}}(f)$ i.e. $\omega(X_f, X_{\varphi_j})$ gives the flow of f along the gauge orbit.

1.7 Geometric Resolution

We have seen that geometrically, it is natural to formulate the theory on the final constraint surface. The reason is that it captures the dynamically accessible points of phase space, and one can specify the dynamics on this surface such that there exist solutions at every point. So let us now consider whether, by formulating the theory on the final constraint surface, we can resolve the issues raised earlier. Recall that at issue is the question of how to reconcile the notion of gauge transformations of states and gauge transformations of solutions. Both Dirac (1964) and Pitts (2014b) take gauge transformations to be determined through the dynamics generated by the Total Hamiltonian, but this leads to different definitions in the case of states and of solutions, and consequently different opinions about whether one should extend the Hamiltonian or not. We can summarize the reasoning common to Dirac (1964) and Pitts (2014b) as follows:

1. First, one determines the gauge transformations using the Total Hamiltonian.
2. Then, one uses the gauge transformations to say whether one should extend the Hamiltonian or not.

I will argue that this reasoning is flawed in three parts. First, I argue that Extended Hamiltonian is motivated independently from consideration of the gauge transformations, and so (2) is wrong: the gauge transformations do not determine the correct form of the Hamiltonian. Second, I argue that the gauge transformations on states arise naturally from the structure of the final constraint surface, without considering the solutions to the equations of motion, and so (1) is wrong: the gauge transformations on states are not simply a special case of the gauge transformations on solutions. Finally, I use these two arguments to show that the gauge transformations on solutions (properly understood) are generated by arbitrary combinations of first-class constraints.

1.7.1 Motivating the Extended Hamiltonian

First, let's start with why the Extended Hamiltonian is motivated. It is clear that on the final constraint surface, Hamiltonians related by an arbitrary combination of first-class constraints are equivocated; they correspond to the same Hamiltonian function on the final constraint surface. This immediately provides one motivation for the Extended Hamiltonian, if one is convinced that we ought to formulate the theory on the final constraint surface. However, what we want to capture is the non-uniqueness of the solutions to the equations of motion. On the final constraint surface, this is captured by the fact that the vector fields corresponding to solutions to the equations of motion for some Hamiltonian are defined only up to arbitrary combinations of vector fields associated with the first-class constraints. Take a (first-class) Hamiltonian vector field X_H and transform it to $X_H + a^j X_{\varphi_j}$ where X_{φ_j} are the null vector fields associated with the first-class constraints φ_j and a^j are arbitrary functions. We have that

$$\tilde{\omega}_f(X_H + a^j X_{\varphi_j}, \cdot) = \tilde{\omega}_f(X_H, \cdot) = dH|_{\Sigma_f}$$

since X_{φ_j} are null vector fields. But this means that transforming X_H by an arbitrary linear combination of the vector fields associated with the first-class constraints preserves the dynamical equations on the final constraint surface. In other words, the structure of the final constraint surface is such that the evolution generated by X_H and that generated by $X_H + a^j X_{\varphi_j}$ is not distinguished: if f satisfies $\tilde{\omega}_f(X_f, X_H) = \frac{df}{dt}|_{\Sigma_f}$, then it satisfies $\tilde{\omega}_f(X_f, X_H + a^j X_{\varphi_j}) = \frac{df}{dt}|_{\Sigma_f}$. Therefore, we can think of the vector fields $X_H + a^j X_{\varphi_j}$ on the final constraint surface as characterizing the equivalence class of vector fields that generate solutions to the equations of motion. Let us call this equivalence class of vector fields the "Extended Hamiltonian vector fields". This provides a second motivation for the Extended Hamiltonian: on the final constraint surface, there is an equivalence class of vector fields associated with the Hamiltonian defined up to the vector fields associated with all of the first-class constraints.

Notice that in such reasoning, we have not made any assumptions about the X_{φ_j} being associated with primary or secondary first-class constraints, nor about what the gauge-transformations are; each first-class constraint constitutes a null direction on the final constraint surface, and it is this property that is important in determining which transformations of the Hamiltonian vector field are dynamically equivalent. In particular, notice that the sense of dynamical equivalence here is just that these Hamiltonian vector fields form an equivalence class, relative to the structure of the constraints surface. Inasmuch as this structure is how one makes predictions in the theory, these Hamiltonian vector fields generate the same predictions.

This provides one argument for why restricting to the Total Hamiltonian is unnatural in the geometric framework: it distinguishes a class of null vectors (those that correspond to primary first-class constraints) that cannot be distinguished from other null vectors in terms of the structure of the final constraint surface.[9]

1.7.2 State Gauge Transformations

Second, let's consider the notion of a gauge transformation on states. Consider the functions that vary along the gauge orbits on Σ_f. These are the functions g for which $\omega(X_g, X_{\varphi_j}) \neq 0$ when restricted to Σ_f. Notice that this means that X_g defined via $\omega(X_g, \cdot) = \mathbf{d}g$ is not tangent to Σ_f. But the induced two-form $\tilde{\omega}_f$ only acts on vector fields that are tangent to Σ_f, and so one cannot write $\tilde{\omega}_f(X_g, \cdot) = \mathbf{d}g|_{\Sigma_f}$. This means that one cannot define the change in g along the gauge orbits using the structure of Σ_f i.e. in terms of $\tilde{\omega}_f$. Therefore,

[9] One can distinguish the secondary constraints through the fact that they correspond to time derivatives of the primary constraints, but this is not the relevant kind of difference in determining the equivalence class of Hamiltonians.

there is a sense in which the structure of the final constraint surface does not distinguish the value of a function at different points along the gauge orbits.

We can make this more precise. Consider the transformation h that takes one along the gauge orbits by an arbitrary amount each point on Σ_f. Then the following is true:

> **Proposition 1.1:** h is an automorphism of the structure $(\Sigma_f, \tilde{\omega}_f)$ i.e. h is a diffeomorphism $h : \Sigma_f \to \Sigma_f$ such that $h^*(\tilde{\omega}_f) = \tilde{\omega}_f$.
>
> *Proof*: Since h takes each point on Σ_f to another arbitrary point along the gauge orbit associated with the first-class constraints φ_j at that point, we can represent h as the flow of the vector field associated with $\alpha^j d\varphi_j$ where α^j are arbitrary functions. Since $d\varphi_j = 0$ on Σ_f, $\alpha^j d\varphi_j = 0$. This means that $\alpha^j d\varphi_j$ is closed i.e. $d(\alpha^j d\varphi_j) = 0$. But this means that one can (locally) associate a vector field Y with $\alpha^j d\varphi_j$ via $\tilde{\omega}_f(Y, \cdot) = \alpha^j d\varphi_j$. It follows that the flow of Y on Σ_f consists of maps that preserve $\tilde{\omega}_f$.[10] So h is a diffeomorphism that takes $\tilde{\omega}_f$ to itself.

Therefore, the structure of the final constraint surface is such that it does not depend on the value that functions take along the gauge orbits; we can move the points along the gauge orbits around arbitrarily and preserve the structure of the the final constraint surface. This provides a precise sense in which the gauge orbits naturally characterize the equivalence classes of states on the final constraint surface.

Notice that this reasoning does not make reference to the dynamics. In particular, it doesn't make reference to the Total Hamiltonian, since it relies only on the structure of the constraint surface. This suggests a revision to the definition of the state gauge transformations:

[10]This follows from Abraham and Marsden (1987) Proposition 3.3.6. (when we extend the proposition to presymplectic manifolds).

State* Gauge Transformation: A (state) gauge transformation is a transformation that relates any two states on the constraint surface that cannot be distinguished by the induced two-form.

This emphasizes that what makes states along a gauge orbit equivalent has to do with their role in the structure of the constraint surface. Notice that on this definition, arbitrary combinations of first-class constraints generate gauge transformations precisely because they give rise to the gauge orbits. We therefore have a definition of the gauge transformations on states that is motivated independently from the gauge transformations on solutions, but which agrees with both sides of the debate about the generators of gauge transformations on states.

We can also use this argument to oppose a claim made by Henneaux and Teitelboim (1994) (one of the standard textbooks on the constrained Hamiltonian formalism). They say:

> "The identification of the gauge orbits with the null surfaces of the induced two-form relies strongly on the postulate made throughout the book that all first-class constraints generate gauge transformations." (p. 54)

In other words, they suggest that one must independently maintain that first-class constraints generate gauge transformations in order to interpret the null surfaces as the gauge-equivalent points. But the argument above shows that this interpretation is motivated from within the geometric formulation.

1.7.3 Solution Gauge Transformations

Finally, let us consider the gauge transformations on solutions. We have determined that on Σ_f, Hamiltonian vector fields that differ by some combination of null vector fields are not distinguished by the structure of Σ_f. The integral curves of these vector fields differ only with regard to where on the gauge orbit they lie at each point in time. Therefore, transforming a solution by an arbitrary amount along the gauge orbit at each point gives rise to another solution generated by a Hamiltonian vector field with a different combination of null vectors. Moreover, we have determined that the states along a gauge orbit form an equivalence class of states. Therefore, we have a natural reason to think that solutions that differ just in terms of where each point lies along the gauge orbit are physically equivalent since the Hamiltonian vector fields that generate these solutions form an equivalence class, and the states along a gauge orbit form an equivalence class.

To see this more precisely: take a curve $s(t)$ defined on the final constraint surface whose tangent vector is a solution to the equations of motion $\tilde{\omega}_f(X_s, X_H + a^j X_{\varphi_j}) = \frac{ds}{dt}$. Now take another curve $s'(t) = h(t) \cdot s(t)$ where $h(t)$ is a smooth function that "moves" $s(t)$ by some arbitrary amount along the gauge orbit at each point. Then $X_{s'}$ will also be a solution to $\tilde{\omega}_f(X_{s'}, X_H + a^j X_{\varphi_j}) = \frac{ds'}{dt}$, since this equation of motion determines the tangent vector to the dynamical trajectory only up to the addition of an arbitrary (time-dependent) combination of null vectors. Therefore, an arbitrary combination of first-class constraints generates a transformation that takes solutions to equivalent solutions on Σ_f.

This motivates the following characterization of the solution gauge transformations:

> **Solution* Gauge Transformation:** A (solution) gauge transformation relates any two curves that are possible evolutions from an initial state under the dynamics generated by the Extended Hamiltonian vector fields.

Notice that this definition is supported on two fronts. First, we have independently motivated the Extended Hamiltonian vector fields as the correct equivalence class. Second, we have independently motivated the equivalence class of states as given by the gauge orbits.

This provides another sense in which restricting to the Total Hamiltonian is unnatural geometrically: it would be to say that the dynamics can distinguish states along a gauge orbit, even though the structure of the final constraint surface is such that it cannot distinguish these states. So we shouldn't think that gauge transformations on states are a special case of those on solutions; rather, they characterize two independent notions that are coherent with each other.

In summary, we can diagnose the conflict between Dirac (1964) and Pitts (2014b) about the correct characterization of a gauge transformation as follows: they both take for granted that the gauge transformations are determined by the evolution generated by the Total Hamiltonian. This leads to a disagreement about the generators of gauge transformations, and consequently the right equivalence class of Hamiltonians. What I have argued here is that this reasoning is flawed: the Extended Hamiltonian can be motivated as the right equivalence class of Hamiltonians before determining the gauge transformations, and the gauge transformations on states can be determined without directly considering the evolution generated by the equivalence class of Hamiltonians. This allows one to maintain a clear conceptual difference between gauge transformations on states and gauge transformations on solutions, and it allows one to maintain that both of these notions capture a notion of physical equivalence without conceptual tension. It also provides support for the position held by Pooley and Wallace (2022): we ought to use the Extended Hamiltonian to determine the gauge transformations on solutions because it is the *correct* equivalence class of Hamiltonians from the perspective of the geometric formulation.

1.8 Possible Counterarguments

Let us now consider how one might respond to the argument given in the previous section; in particular, how one might defend "Solution Gauge Transformation" over "Solution*

Gauge Transformation", since it is these notions that lead to different characterizations of the transformations that generate gauge transformations. First, I will consider a recent argument given by Pitts in response to Pooley and Wallace (2022) that the Extended Hamiltonian is a trivial reformulation of a theory, and therefore that it is not a physically interesting alternative. Then, I will turn to an objection that we should not commit to the geometry of the constraint surface as a guide to the symmetries of the theory.

1.8.1 Triviality Argument

Given Pooley and Wallace's response to Pitts (2014b), there seem to be two possible reactions. First, one could maintain that the question of whether arbitrary combinations of first-class constraints generate gauge transformations or not comes down to whether the Total or Extended Hamiltonian is considered the right equivalence class of Hamiltonians. Second, one could maintain that what the debate shows is that there is no Hamiltonian-independent way to characterize the gauge transformations but that we can think of these different forms of the Hamiltonian as equivalent and so there is no conflict. This second reaction is the approach that Pitts (2022, 2024) takes: he argues that the Extended Hamiltonian can be seen as a trivial kind of reformulation of the theory. Moreover, he argues that we can construct similar kinds of reformulation that allow one to conclude that quantities other than the first-class constraints generate gauge transformations, which is not maintained by either side of the debate. So we haven't gained insight into the gauge transformations by making this move; rather, we have stated the same thing in a different (more complicated) form.

To see what Pitts means by a trivial reformulation, let us consider a simple example presented in Pitts (2022). Take a Lagrangian given by:

$$L = \frac{1}{2}\dot{q}^2 \tag{1.14}$$

This describes a particle moving in a straight line with uniform velocity; the equation of motion is $\frac{d^2q}{dt^2} = 0$. The symmetries of this equation of motion are spatial translations and boosts. Now consider the Lagrangian:

$$L = \frac{1}{2}(\dot{q} - \dot{\mu})^2 \tag{1.15}$$

where μ is either an arbitrary function of time or a dynamical variable. This Lagrangian is invariant under the transformation $q \to q + \epsilon$, $\mu \to \mu + \epsilon$ where ϵ is an arbitrary function of time. So, Pitts argues, we have 'added' a symmetry. Moreover, this Lagrangian gives rise to the equation of motion $\frac{d^2(q-\mu)}{dt^2} = 0$, which says that $q - \mu$ represents a particle moving in a straight line with uniform velocity. Therefore, the new Lagrangian has the same physical content as the previous Lagrangian.

More precisely, we can see in the Hamiltonian formulation that the first Lagrangian does not have any gauge freedom since there are no constraints. However, treating μ as a dynamical variable, we find that the Hamiltonian formulation of the second Lagrangian has a constraint, namely $p_q + p_\mu$, which is first-class. So, there is a new gauge transformation in the Hamiltonian formulation: it is the transformation $q \to q + \epsilon$, $\mu \to \mu + \epsilon$ generated by the first-class constraint.

Pitts calls this process of revising a Lagrangian by adding a new variable (or "splitting one quantity into two") and thereby adding new symmetries "de-Ockhamization". In the above sense, Pitts argues it is trivial: it doesn't change the physical content of the theory, and therefore it is just a more complex redefinition of the original theory.

To further push this point, Pitts shows that we can do the same thing to reach the conclusion that second-class constraints generate gauge transformations, which is arguably a *reductio ad absurdum*. Take the Lagrangian for Electromagnetism, but add a photon mass term $-\frac{1}{2}m^2(\vec{A}^2 - V^2)$. This is called "Proca Electromagnetism". The primary constraint is the same, but the secondary constraint has an additional term of m^2V. This has the consequence that both constraints are second-class: $\{p_V, \nabla \cdot p_{\vec{A}} + m^2V - \rho\} = m^2$.

The time derivative of the secondary constraint fixes the value of λ: one gets $\lambda = \nabla \cdot \vec{A}$. Therefore, one can remove the arbitrariness in the Total Hamiltonian. In particular, the primary second-class constraint generates a transformation that takes $V \to V + \alpha$ where α is an arbitrary function of time, as in ordinary Electromagnetism, but it does not generate a gauge transformation. Consider the equations of motion for p_v:

$$\frac{\partial p_V}{\partial t} = \nabla \cdot p_{\vec{A}} - \rho + m^2V \tag{1.16}$$

The right-hand side is just the secondary constraint and so is equal to 0. But if we transform $V \to V + \alpha$, the right-hand side is equal to $m^2\alpha \neq 0$.

However, let's consider "de-Ockhamizing" this theory by replacing V with $V + \mu$ in the equations of motion where μ is arbitrary. Now the above equation of motion is satisfied when we transform $V \to V + \alpha$, since the right-hand side is just equal to an arbitrary function. Moreover, we can think of the de-Ockhamized equations of motion as resulting from an "extended Hamiltonian", where V is replaced by $V + \mu$.

Although extending the Hamiltonian in this way isn't to add a linear combination of constraints, Pitts suggests that it is analogous to what Pooley and Wallace (2022) do in order to recover the claim that arbitrary combinations of first-class constraints generate gauge transformations in Electromagnetism: one redefines a quantity by adding a new variable. In doing so, one introduces new symmetries, but these should not be regarded as "genuine" gauge transformations; if they were, then one would have to conclude that second-class constraints generate genuine gauge transformations as well. Therefore, Pitts concludes, we should not think that the Extended Hamiltonian supports the claim that gauge transformations are generated by arbitrary combinations of first-class constraints in a non-trivial sense.

Response to the Triviality Argument

Let us look more closely at the first example presented by Pitts. Recall that the claim is that the Lagrangian $L = \frac{1}{2}(\dot{q} - \dot{\mu})^2$ has *more* symmetries than the Lagrangian $L = \frac{1}{2}\dot{q}^2$ even though they have the same empirical content. The sense in which it has more symmetries is supposed to be that if we just consider the variable q, then for the original Lagrangian, we can only transform q by spatial translations and boosts and preserve the equations of motion, while for the "de-Ockhamized" Lagrangian, we can transform q by an arbitrary function of time and preserve the equations of motion (with a corresponding change to μ).

There are two kinds of comparison here: First, there is a comparison regarding *empirical content*. Second, there is a comparison regarding *symmetries*. In order to maintain simultaneously that the Lagrangians are empirically equivalent and that they have different symmetries, it must be that we are comparing the Lagrangians in the same way when we make this claim. So, let us consider under what standard of comparison one can make these claims.

Starting with the claim that the two Lagrangians are empirically equivalent, this seems to rely on taking q to represent the position of the particle in the first Lagrangian and $q - \mu$ to represent the position of the particle in the second Lagrangian, since these quantities satisfy the same equations of motion, namely that the second derivative is equal to 0.

However, if we identify q in the first Lagrangian with $q - \mu$ in the second Lagrangian, then we also should compare the transformations of q that preserve the E-L equations for the first Lagrangian with the transformations of $q - \mu$ that preserve the E-L equations for the second Lagrangian. But these transformations are the same: the only transformations of $q - \mu$ that preserve the equations of motion for the second Lagrangian are spatial translations and boosts. Indeed, all we have done is effectively change the label of the variable that represents position. This is clearly a trivial kind of reformulation. However, it does not support Pitts' position that the second Lagrangian has additional gauge symmetries, since under the standard of comparison where q is identified with $q - \mu$, the Lagrangians are empirically equivalent and also have the same symmetries.

One might try to respond by saying the following: the transformation $q \to q + \epsilon$, $\mu \to \mu + \epsilon$, where ϵ is an arbitrary function of time, is a symmetry of the second Lagrangian that is not a symmetry of the first Lagrangian and that preserves the same form of the equations of motion. But under the identification of q with $q - \mu$, this transformation *is* a symmetry of the first Lagrangian – it is the identity transformation on q. If instead, one said that the transformation $q - \mu \to q - \mu + \epsilon$ is a symmetry of the second Lagrangian by taking μ to be arbitrary, then this would be an 'additional' symmetry, but it would also mean that the Lagrangians are not empirically equivalent via identification of q with $q - \mu$; one describes a particle moving in a straight line and the other describes a particle whose dynamics is arbitrary. Either way, one cannot simultaneously claim that one has added a new symmetry and preserved the empirical content under the identification of q in the first Lagrangian with $q - \mu$ in the second Lagrangian.

Another way to compare the Lagrangians is to identify the quantity q as representing the position of the particle in both Lagrangians and take μ in the second Lagrangian to represent an additional (perhaps arbitrary) variable. On this standard of comparison, there is a sense in which one has added a symmetry of q by moving to the second Lagrangian: we can transform q by an arbitrary function of time and preserve the equations of motion. But now, we have that the equation of motion for q is: $\frac{d^2q}{dt^2} = \frac{d^2\mu}{dt^2}$. This is a different equation of motion for q compared to the original Lagrangian since it describes a situation where the position of the particle is either an arbitrary function of time, when μ is arbitrary, or where the particle moves in the same way as μ, when μ is a dynamical variable. Therefore, under this standard of comparison, the two Lagrangians are not empirically equivalent, and so it is not a trivial reformulation.

The upshot is that one cannot simultaneously maintain that the two Lagrangians are empirically equivalent and that one has more symmetries than the other. The same is true of the second example in Section 1.8.1 of Proca Electromagnetism. For $V \to V + \alpha$ to be a gauge transformation, V must be arbitrary in the equations of motion. But V is not arbitrary in the original equations of motion; the equations of motion for V are:

$$\frac{\partial V}{\partial t} = \lambda$$

where λ satisfies $\lambda = \nabla \cdot \vec{A}$, and so is not arbitrary. Therefore, if the de-Ockhamization involves replacing V with an arbitrary function of time $\mu' = V + \mu$, the two equations of motion are not empirically equivalent.[11] If instead one wants to maintain that $V + \mu$ plays the same role as V in the original equation of motion, then the equations would be empirically equivalent but they would also have the same symmetries: the transformation

[11] Indeed, if we replace V everywhere in the Total Hamiltonian with an arbitrary function μ', then we would not have any secondary constraints and there would be one first-class constraint that generates a gauge transformation that shifts V by an arbitrary function of time, as one would expect if V is decoupled from the equations of motion. This would be a different theory from Proca Electromagnetism.

$V + \mu \to V + \mu + \alpha$ would not be a symmetry if $V + \mu$ is understood to be a relabelling of V.

Consequently, there is a kind of trivial reformulation that one can invoke in the examples that Pitts provides, but it isn't the kind where we add new symmetries. Indeed, we should not be surprised that changing the symmetries of the theory in general has new empirical consequences: symmetries tell us which physical situations are equivalent according to a theory, and so theories with different symmetries will disagree about the physical possibilities.

Is the Extended Hamiltonian Trivial?

We have established that there is a kind of reformulation of a theory that is trivial but does not correspond to adding new symmetries to a theory. On the other hand, there is a kind of reformulation that does change the symmetries of a theory, but in the examples that Pitts gives, this reformulation leads to a Lagrangian that is physically distinct. So the natural question is: What kind of reformulation is going on in the case where one moves from the Total Hamiltonian to the Extended Hamiltonian?

Let us consider this question in the context of classical Electromagnetism. Again, we need to consider what the standard of comparison is supposed to be. If the replacement of V with $V + \mu$ corresponds to a mere relabeling, then the move to the Extended Hamiltonian appears trivial. But this would mean that we understand V in the Total Hamiltonian and $V + \mu$ in the Extended Hamiltonian to have the same symmetries. This does not seem to be what is going on; the Extended Hamiltonian is supposed to come with the addition of new symmetries, namely, the transformations generated by an arbitrary combination of the first-class constraints (that go past the transformations generated by the specific combination of first-class constraints). This suggests that the move to the Extended Hamiltonian corresponds to the second kind of reformulation: one that changes the physical content of the theory. But then, how is this compatible with the claim of Pooley and Wallace (2022) that the Extended Hamiltonian doesn't come with a change in empirical content?

I think that the geometric formulation presented in Section 1.6 allows us to see what is going on. Recall that the Total Hamiltonian is the equivalence class of Hamiltonians defined up to arbitrary combinations of primary (first-class) constraints. These Hamiltonians are equivocated when we move to the submanifold of T^*Q defined by the satisfaction of the primary constraints. We call this submanifold the *primary constraint surface*. Therefore, we can think of the theory described by the Total Hamiltonian geometrically as the object $(\Sigma_p, \tilde{\omega}_p, H, \varphi_i)$ where Σ_p is the primary constraint surface, $\tilde{\omega}_p$ is the presymplectic two-form defined intrinsically on the primary constraint surface, H is the Hamiltonian restricted to the primary constraint surface, and φ_i are the secondary constraints. The dynamics is given by two equations, $\tilde{\omega}_p(X_H, \cdot) = \mathbf{d}H$ and $\varphi_i = 0$. Notice that on the primary constraint surface, the solution to $\tilde{\omega}_p(X_H, \cdot) = \mathbf{d}H$ is unique only up to arbitrary combinations of vector fields associated with the primary first-class constraints (the null vector fields of $\tilde{\omega}_p$), which characterizes the sense in which there are multiple, equivalent solutions to the equation of motion using the Total Hamiltonian.

Similarly, the Extended Hamiltonian is the equivalence class of Hamiltonians defined up to arbitrary combinations of all the first-class constraints, which are equivocated on the surface defined by the satisfaction of the primary and secondary constraints. We call this the *final constraint surface*. Therefore, we can think of the theory described by the Extended Hamiltonian as the object $(\Sigma_f, \tilde{\omega}_f, H)$ where Σ_f is the final constraint surface, $\tilde{\omega}_f$ is the presymplectic two-form defined intrinsically on the final constraint surface, and H is Hamiltonian restricted to the final constraint surface. The solution to $\tilde{\omega}_f(X_H, \cdot) = \mathbf{d}H$ on the final constraint surface is unique only up to arbitrary combinations of vector fields associated with the first-class constraints.

This characterization naturally provides a sense in which the Extended Hamiltonian theory regards more solutions as equivalent compared to the Total Hamiltonian theory: there are solutions that the Total Hamiltonian theory distinguishes between that the Extended Hamiltonian theory does not distinguish between (when we consider these solutions on the final constraint surface). More generally, we can now give a precise characterization of the fact that the Extended Hamiltonian theory has more symmetries than the Total Hamiltonian theory: the null vector fields of the two-form on the primary constraint surface are a subset of the null vector fields of the two-form on the final constraint surface, and so the gauge transformations – the transformations along the gauge orbits – of the Total Hamiltonian theory are a subset of the gauge transformations of the Extended Hamiltonian theory (when considered on the final constraint surface).

However, restriction to the final constraint surface does not come with a change in empirical content. One way to put the reason for this is that the solutions to the equations of motion on the final constraint surface correspond precisely to the solutions to the equations of motion on the unconstrained symplectic manifold that satisfy the constraints (equivalently, the solutions to the equations of motion on the primary constraint surface that satisfy the secondary constraints); they are the solutions projected to the final constraint surface. And the symmetries of *these* solutions, i.e. the solutions defined on the final constraint surface, are just the symmetries given by the Extended Hamiltonian theory. Therefore, as long as constraints are considered a physical requirement, the Extended Hamiltonian theory is empirically equivalent to the Total Hamiltonian theory.

Indeed, I think that the geometric framing helps to see exactly what the move to the Extended Hamiltonian formalism corresponds to: it corresponds to moving to a theory with *less structure*, since it is a theory with more symmetries, that nonetheless has the same empirical content. In other words, it shows that the theory defined by the Total Hamiltonian has a kind of 'excess structure': there are points and solutions distinguished by the theory formulated on the primary constraint surface that are not distinguished by a theory that maintains the same physical content and yet has less structure. In particular, the points that lie along the integral curves of the vector fields associated with the secondary first-class constraints are symmetry-related in the Extended Hamiltonian formalism, but not in the Total Hamiltonian formalism. We have reason to think that they should be symmetry-related precisely because the differences between these points do not seem to be playing any role in the empirical content of the theory.

To spell out the sense in which the Extended Hamiltonian formalism has less structure than the Total Hamiltonian formalism more precisely, let us define the theories in category-theoretic terms.[12] Take the category **TotHam** to have as objects the models $(\Sigma_p, \tilde{\omega}_p, H, \varphi_i)$, and let us take the arrows between objects $(\Sigma_p, \tilde{\omega}_p, H, \varphi_i)$, $(\Sigma_p, \tilde{\omega}'_p, H', \varphi'_i)$ to be the diffeomorphisms $f : \Sigma_p \to \Sigma_p$ such that $f^*(\tilde{\omega}'_p) = \tilde{\omega}_p$, $f^*(H') = H$ and $f^*(\varphi'_i) = \varphi_i$ i.e. the symmetries are taken to be the symplectomorphisms that preserve the Hamiltonian and the secondary constraints. Similarly, let us take the category **ExtHam** to have as objects the models $(\Sigma_f, \tilde{\omega}_f, H)$ and as arrows between objects $(\Sigma_f, \tilde{\omega}_f, H)$, $(\Sigma_f, \tilde{\omega}'_f, H')$ the diffeomorphisms $g : \Sigma_f \to \Sigma_f$ such that $g^*(\tilde{\omega}'_f) = \tilde{\omega}_f$ and $g^*(H') = H$.

Relations between theories are described by functors between the categories representing those theories. A functor $F : \mathcal{C} \to \mathcal{D}$ from the category $\mathcal{C}$ to the category $\mathcal{D}$ is said to be *full* if for every pair of objects A, B of $\mathcal{C}$ the map $F : \text{hom}(A, B) \to \text{hom}(F(A), F(B))$ induced by F is surjective, where $\text{hom}(A, B)$ is the collection of arrows from A to B. Similarly, F is said to be *faithful* if for every pair of objects the induced map on arrows is injective. Finally, F is said to be *essentially surjective* if for every object X of $\mathcal{D}$, there is some object A of $\mathcal{C}$ such that $F(A)$ is isomorphic to X. Using this terminology, we say (following Weatherall (2016b)) that a theory represented by category $\mathcal{C}$ has more structure than a theory represented by

[12] Category theory has been used in several places to give a precise sense in which one theory has less structure than another. See, for example, Weatherall (2016b); Nguyen et al. (2020); Bradley and Weatherall (2020). For a defense of using category theory to represent theories more generally, see Halvorson (2012, 2016); Halvorson and Tsementzis (2017); Weatherall (2016a, 2017).

category $\mathcal{D}$ if a functor $F : \mathcal{C} \to \mathcal{D}$ is not full (but is faithful and essentially surjective). In this case, we say that F *forgets (only) structure.*[13]

Consider the functor F : **TotHam** $\to$ **ExtHam** that takes each model $(\Sigma_p, \tilde{\omega}_p, H, \varphi_i)$ to its restriction to the points that satisfy the constraints $\varphi_i = 0$, i.e. the associated model $(\Sigma_f, \tilde{\omega}_f, H)$, and that takes the arrow f to its action on Σ_f (since f preserves the secondary constraints, f preserves Σ_f, and so this is well-defined). Then:

Proposition 1.2: F : **TotHam** $\to$ **ExtHam** forgets (only) structure.[14]

Inasmuch as forgetting structure in category-theoretic terms captures what it is for one theory to have less structure than another, this proposition gives a precise characterization of the sense in which the Extended Hamiltonian formalism has less structure than the Total Hamiltonian formalism. Therefore, contra Pitts, the move to the Extended Hamiltonian should not be thought of as simply redefining a theory in terms of new quantities. Rather, it defines a new theory that removes structure from the Total Hamiltonian theory. This means that there is a genuine disagreement about the gauge transformations between the two theories. But even further, it suggests that the standard view that arbitrary combinations of first-class constraints is correct, by Pitts' own lights: moving to the Extended Hamiltonian is the opposite of "de-Ockhamization", in the sense that it is a simpler theory (in terms of structure) than the Total Hamiltonian theory. So Pitts was right that simplicity considerations matter in the debate between the Total and Extended Hamiltonian, but rather than these considerations pushing one towards the Total Hamiltonian, they push one towards the theory that captures all of the symmetries that the theory is naturally understood as having: the theory characterized by the Extended Hamiltonian.

[13]The terminology of forgetful functors originates with Baez et al. (2004).
[14]See A.1 for proof.

1.8.2 Against the Constraint Surface

Let's consider two objections one might have to taking the geometry of the constraint surface as a guide to the symmetries of the theory: First, the objection that we shouldn't restrict to the constraint surface. Second, the objection that we shouldn't think that the geometrical formulation of the constrained Hamiltonian formalism is adequate.

Let's start with the first objection. Inasmuch as constraints are understood to provide the "physically allowed states", it seems natural to think that the points off the constraint surface are unnecessary for describing the dynamics of the theory. However, one might want to maintain that these points still have importance as "kinematically possible" states. That is, one might want to maintain that we ought to consider states off the constraint surface as important for describing the physical theory as a whole, even if the dynamics is restricted to the constraint surface. In particular, the secondary constraints are fixed by thinking about the consistency of the primary constraints *with the dynamics*. And so it might seem that at least when it comes to secondary constraints, there is no logical inconsistency with specifying a theory in terms of points where the secondary constraints do not hold. And the vector fields associated with the secondary constraints are not null vectors of the two-form on the full phase space (nor on the primary constraint surface); the full phase space is symplectic, and so it is non-degenerate by definition. So, the counterargument goes, we cannot use the fact that these vector fields are null to argue that they generate gauge transformations.

One natural response is that the points off of the constraint surface are a kind of 'excess structure': although there is nothing inconsistent about including them, the content of the theory is given by the constraint surface. We already gave a response of this kind in Section 1.8.1. However, I think another response is to point out that the idea that we start with the primary constraints and then generate the secondary constraints through the dynamics is somewhat an accident of the way that the Hamiltonian formalism is usually set up. As I presented Dirac's version of the theory, one starts with a Lagrangian function, from which one derives the primary constraints. Only once we have the primary constraints and the Hamiltonian in hand do we determine the secondary constraints. But we could have set up the Hamiltonian formalism in a different way: we could say that our theory is given by specifying a Hamiltonian function, a symplectic two-form, and a collection of constraints. In this way of setting up the formalism, although there is a functional relationship between the primary and secondary constraints, there is no clear difference in the role that they play. In particular, the only relevant difference seems to be which constraints are first-class; these are the ones that generate transformations that keep one along the constraint surface, and which are null vectors of the induced two-form on the constraint surface.

In order to push back on this response, one would have to argue that there is something wrong with setting up the Hamiltonian formalism in this way. This leads to the second objection: that the geometric formulation of the Hamiltonian formalism is not adequate. The first thing to note here is that the geometric formulation is a natural extension of a widely accepted formulation of Hamiltonian mechanics without constraints using symplectic manifolds. That is, this formulation takes the standard geometric way of expressing Hamiltonian mechanics and considers what changes when one adds constraints, and so in this sense is well motivated. But one might want to argue that it is inadequate in a different way. In particular, one might want to argue that the Hamiltonian formalism is necessarily derivative from the Lagrangian formalism; the Lagrangian formalism is the "fundamental" one, and the Hamiltonian formalism is just an alternative way of expressing this formalism. On this view, there is a difference between the primary and secondary constraints that originates with the Lagrangian viewpoint and that isn't captured purely by considering the geometry of the Hamiltonian formalism. The difference is that the primary constraints are necessary to ensure that the Hamiltonian formalism is equivalent to the Lagrangian formalism, while the secondary constraints are 'extra' constraints on the Hamiltonian side that are not motivated from the Lagrangian perspective. In particular, it is only the primary constraints that are imposed in order for the map from the Lagrangian to Hamiltonian state spaces to be invertible.[15]

Therefore, this argument goes, restricting to the secondary constraint surface – and consequently having the view that arbitrary combinations of first-class constraints generate gauge transformations – leads to a theory that is inequivalent to the Lagrangian theory, and so is not the right theory to consider. Indeed, one can show that the Total Hamiltonian formalism, understood as relying on the primary constraint surface, gives rise to solutions that are equivalent to the solutions to the Euler-Lagrange equations (Batlle et al. (1986)). Therefore, it seems that restricting to the constraint surface (including the secondary constraints) gives rise to a theory that is *empirically* equivalent to the Lagrangian formalism but is not strictly the same. And so, if one takes the view that the Lagrangian formalism is more fundamental, this might motivate one to say that our definition of a gauge transformation should be inherited from this formalism, and thus not the definition motivated by the geometry of the constraint surface.

Responding to this argument requires us to consider deep and subtle questions about what makes one theory more "fundamental" than another and how to characterize the equivalence of theories. This will be the task of Chapter 2.

1.9 Conclusion

To summarize, I have argued that the debate about the correct characterization of the gauge transformations in the constrained Hamiltonian formalism rests on assumptions about the relationship between gauge transformations and the form of the Hamiltonian that are unnatural from the perspective of the geometric formulation of the constrained Hamiltonian

[15]The transformation taking points $(x, \dot{x})$ to $(x, \frac{\partial L}{\partial \dot{x}})$ is called the Legendre transformation, and in the case where it is non-invertible, its image is the primary constraint surface.

formalism. Using the geometric formulation, I showed that we can distinguish between gauge transformations on states and gauge transformations on solutions in a conceptually clear way and that both are generated by arbitrary combinations of first-class constraints, thereby supporting the orthodox view. However, this allowed us to pinpoint more clearly where disagreement can be found. In particular, I suggested that there are crucial questions about the relationship between Lagrangian and Hamiltonian theories in the presence of gauge symmetry, where different answers to these questions can lead to different views regarding the correct form of the Hamiltonian, and thus to what the correct characterization of the gauge transformations is.

One important topic that I have not discussed in this chapter is the “Problem of Time". Recall: for theories that are time-reparameterization invariant, the standard account of gauge transformations implies that time evolution is itself a gauge transformation since the Hamiltonian is a first-class constraint. In supporting the standard account of gauge transformations as being generated by arbitrary combinations of first-class constraints, it might appear that we are also left with the issues surrounding the Problem of Time. That is, we haven't seemed to do anything to deny that a Hamiltonian first-class constraint generates a gauge transformation. However, I think that the distinctions drawn out here highlight what is interesting about the case of a Hamiltonian first-class constraint. In particular, the claim that we can conceptually distinguish the gauge transformations on states and the gauge transformations on solutions does not seem to be true in the case where the Hamiltonian is a first-class constraint: the gauge orbits are just the solutions to the equations of motion, and so the states along a gauge orbit cannot be understood independently from the dynamics. Thus, it is less clear whether one can distinguish two notions of physical equivalence as well. This suggests that the puzzle surrounding the Problem of Time at least partly comes down to the fact the transformation generated by a Hamiltonian first-class constraint doesn't fall neatly into the categories defined here. But more work needs to be done to say what exactly is distinct about this case, inasmuch as when the Hamiltonian is a first-class constraint, it seems to play the same role geometrically as any other first-class constraint since it is a null vector of the induced two-form on the constraint surface. To answer this would require a more careful consideration of the role of the Hamiltonian and whether there is a more fine-grained distinction between different kinds of constraints. I hope that the work here has at least provided support for the claim that the Problem of Time is not the result of an incorrect definition of the gauge transformations in the constrained Hamiltonian formalism; rather, it must be treated on its own terms.

Chapter 2

The Relationship between Lagrangian and Hamiltonian Mechanics: The Irregular Case

2.1 Introduction

Lagrangian and Hamiltonian mechanics are widely held to be two distinct but equivalent ways of formulating classical theories. Although some philosophers have recently challenged this view[1], Barrett (2019) makes precise the sense in which one can maintain that Lagrangian and Hamiltonian mechanics are equivalent: as long as one characterizes the structure of these theories in a certain natural way, one can show that they are theoretically equivalent, where the standard of theoretical equivalence is categorical equivalence.

However, Barrett's equivalence result is restricted in an important way: he only shows equivalence between "hyperregular" models of Lagrangian and Hamiltonian mechanics. While hy-

[1] See in particular, North (2009) and Curiel (2014).

perregularity characterizes a large class of theories, it does not characterize the class of *gauge theories*: theories that have local, time-dependent symmetries. The question of whether Lagrangian and Hamiltonian mechanics are equivalent in the context of gauge theories is one that has not been discussed directly in the philosophical literature, despite the fact that it bears on other debates that are prominent in the literature. In particular, there has been a recent debate about the correct characterization of the gauge transformations in the Hamiltonian formalism. Several authors have criticized the standard view on the basis that the resulting theory is inequivalent to the Lagrangian formalism.[2] However, one fails to find a clear exposition of which formulations of Lagrangian and Hamiltonian mechanics in the presence of gauge symmetries are equivalent, and in what sense.

In this chapter, I aim to fill this gap. I demonstrate that the relationship between Lagrangian and Hamiltonian mechanics is made significantly more complicated when the assumption of hyperregularity is dropped, and that claims that are made about equivalence in the literature have so far failed to establish more than a notion of *dynamical equivalence* in the non-hyperregular context. However, I show that one can extend Barrett's result to prove an equivalence result in the irregular case by constructing hyperregular models of Lagrangian and Hamiltonian gauge theories through a process known as 'symplectic reduction'. In doing so, I argue that the claims in the literature that the standard approach to gauge transformations renders Hamiltonian mechanics inequivalent to Lagrangain mechanics are false: there is a natural formulation of Lagrangian mechanics in the irregular context that is equivalent to the formulation of Hamiltonian mechanics under the standard definition of gauge transformations.

While ultimately the chapter supports the equivalence between Lagrangian and Hamiltonian mechanics in the context of gauge theories, exploring this question will highlight several interesting questions about the way that one can construct models of Lagrangian mechanics

[2] See in particular Pitts (2014b,a); Gracia and Pons (1988).

from models of Hamiltonian mechanics and vice versa, about the role that *constraints* play in relating the kinematics and dynamics of a theory, as well as the interpretation of gauge transformations.

In Section 2.1, I spell out the equivalence result in Barrett (2019), paying particular attention to the parts of the result that require the assumption of hyperregularity. In Section 2.2, I discuss how the situation changes when one considers gauge theories, and present the standard Hamiltonian approach to determining the gauge transformations in terms of a constraint formalism. In Section 2.3, I consider the arguments in the literature regarding equivalence between Lagrangian and Hamiltonian gauge theories, and I discuss why they fall short of providing an account of *theoretical* equivalence. In Sections 2.4 and 2.6, I show that one can reformulate Lagrangian mechanics as a constraint theory in a way that is analogous to formulating a Hamiltonian constraint theory, drawing from the work of Gotay and Nester (1979), and I show that the models of the reformulated Lagrangian gauge theory are related to the models of the Hamiltonian constraint theory in a natural way. In Section 2.7, I prove an equivalence result that extends the result in Barrett (2019) to the context of gauge theories. Finally, in Section 2.8 I discuss the upshots of this equivalence result and some possible responses.

2.2 The Regular Case

The relationship between Lagrangian and Hamiltonian mechanics in the 'regular' case has been widely discussed. On the one hand, North (2009) defends the view that Hamiltonian mechanics has less structure than Lagrangian mechanics. On the other hand, Curiel (2014) agrees that Hamiltonian and Lagrangian mechanics ascribe different structure, but argues that Lagrangian mechanics is a better representation of the structure of classical systems. More recently, Barrett (2019) argues that this debate hinges on how one defines the structure of the two theories: while one can maintain that they are inequivalent by defining the structure of the two theories in a particular way, there is also a natural way of spelling out the structure of the two theories that renders them equivalent under a widely defended account of theoretical equivalence, namely, categorical equivalence.[3]

In light of this debate, let us distinguish three views that one might hold regarding the equivalence between Lagrangian and Hamiltonian mechanics in the 'regular' case:

Lagrangian-first View: Lagrangian mechanics better represents physical systems than Hamiltonian mechanics.

Hamiltonian-first View: Hamiltonian mechanics better represents physical systems than Lagrangian mechanics.

Equivalence View: Lagrangian and Hamiltonian mechanics are (categorically) equivalent, and so equally well represent physical systems.

Our focus here will be whether the Equivalence View can also be maintained in the irregular case, and so it will be important for our purposes to see how the Equivalence View is defended in the regular case.

Lagrangian mechanics has state space given by the tangent bundle of configuration space, T_*Q, whose points consist of the pair $(q_i, \dot{q}_i)$ encoding the positions and velocities of the particles. The dynamics are given by specifying a Lagrangian function L, with dynamical equations given by the Euler-Lagrange equations, which in coordinate-dependent form are given by: $\dfrac{d}{dt}\dfrac{\partial L}{\partial \dot{q}} = \dfrac{\partial L}{\partial q}$.

The fiber derivative of L is called the Legendre transformation and it is the map FL : $T_*Q \to T^*Q$ from the tangent to cotangent bundle that is defined as taking the point $(q_i, \dot{q}_i)$

[3] See Halvorson (2012, 2016), Weatherall (2016a,b, 2019b) for discussion of categorical equivalence as the standard for theoretical equivalence.

to $(q_i, \frac{\partial L}{\partial \dot{q}_i})$. We say that L is *regular* if FL is a local diffeomorphism. When FL is a global diffeomorphism i.e. it is also invertible, we say that the model (T_*Q, L) is *hyperregular.*

Hamiltonian mechanics has as its state space the cotangent bundle of configuration space, T^*Q, whose points consist of the pair (q_i, p_i) encoding the positions and canonical momenta of the particles. The dynamics are given by specifying a Hamiltonian function H, with dynamical equations given by Hamilton's equations: $\dfrac{dq}{dt} = \dfrac{\partial H}{\partial p}, \dfrac{dp}{dt} = -\dfrac{\partial H}{\partial q}$.

The fiber derivative of H is the map $FH : T^*Q \to T_*Q$ from the cotangent to tangent bundle that is defined as taking the point (q_i, p_i) to $(q_i, \frac{\partial H}{\partial p_i})$. When FH is a (global) diffeomorphism, we say that the model (T^*Q, H) is (hyper)regular.

The cotangent bundle naturally comes equipped with a symplectic (closed, non-degenerate) two-form ω. We can write the equations of motion in terms of this two-form: $\omega(X_H, \cdot) = dH$ where X_H is the vector field associated with the Hamiltonian, which is unique by the non-degeneracy of the symplectic two-form. The integral curves of X_H correspond to solutions.

We can also use this symplectic structure to define a two-form on the tangent bundle, $\Omega = FL^*(\omega)$. Ω is symplectic when FL is a (local or global) diffeomorphism. We can then show that the Euler-Lagrange equations are equivalent to $\Omega(X_E, \cdot) = dE$ where X_E is the vector field associated with the energy function $E = FL(\dot{q}_i)\dot{q}^i - L$. The integral curves of X_E correspond to solutions.

The structure-preserving maps of tangent space are given by point$_*$ transformations T_*f, defined as follows: given a diffeomorphism $f : M_1 \to M_2$, $T_*f : (q, v) \to (f(q), f_*(v))$. Similarly, the structure-preserving maps on cotangent space are given by point* transformations: given a diffeomorphism $f : M_1 \to M_2$, $T^*f : (q, p) \to (f^{-1}(q), f^*(p))$.

Let us restrict ourselves to hyperregular models of Lagrangian and Hamiltonian mechanics. Define the functor F between a hyperregular model of Lagrangian mechanics and a hyperregular model of Hamiltonian mechanics as $F : (T_*Q, L) \rightarrow (T^*Q, E \circ FL^{-1})$, $F : T_*f \rightarrow T^*(f^{-1})$.

Similarly, define the functor G between a hyperregular model of Hamiltonian mechanics and a hyperregular model of Lagrangian mechanics as $G : (T^*Q, H) \rightarrow (TQ, (\theta_a(X_H)^a - H) \circ FH^{-1})$, $G : T^*f \rightarrow T_*(f^{-1})$ where θ_a is the canonical one-form such that $\omega_{ab} = -d_a\theta_b$. These translation maps preserve empirical content, in the sense that they preserve the base integral curves.[4]

Define the categories **Lag** and **Ham** in the following way:

1. An object in the category **Lag** is a hyperregular model (T_*Q, L). An arrow $(T_*Q_1, L_1) \rightarrow (T_*Q_2, L_2)$ is a point$_*$ transformation $T_*f : T_*Q_1 \rightarrow T_*Q_2$ that preserves the Lagrangian in the sense that $L_2 \circ T_*f = L_1$.

2. An object in the category **Ham** is a hyperregular model (T^*Q, H). An arrow $(T^*Q_1, H_1) \rightarrow (T^*Q_2, H_2)$ is a point* transformation $T^*f : T^*Q_1 \rightarrow T^*Q_2$ that preserves the Hamiltonian in the sense that $H_2 \circ T^*f = H_1$.

Theorem (Barrett (2019)): F : **Lag** $\rightarrow$ **Ham** and G : **Ham** $\rightarrow$ **Lag** are equivalences that preserve solutions.

The upshot is that as long as one is concerned with *hyperregular* Lagrangian and Hamiltonian models, there is a clear sense in which these theories are equivalent in terms of categorical equivalence. Indeed, the proof of the above theorem relies on hyperregularity in several ways. First, notice that the functors F and G rely on the maps FL^{-1} and FH^{-1} in order to define a Hamiltonian model in terms of a Lagrangian model and vice versa. These maps are only well-defined functions (globally) if FL and FH are (global) diffeomorphisms. Second,

[4]See Abraham and Marsden (1987) for more details.

Barrett proves the above theorem by showing that F and G are inverses in the sense that $GF(T_*Q, L) = (T_*Q, L)$ and $FG(T^*Q, H) = (T^*Q, H)$ (and similarly are inverses on arrows). This relies on the fact that $FL^{-1} = FH$ and $FH^{-1} = FL$, which is only true in the hyperregular context.

Given the importance of hyperregularity in reaching the conclusion that the categories of Lagrangian and Hamiltonian models are equivalent, one might conclude that the class of irregular Lagrangian and Hamiltonian theories cannot be categorically equivalent.[5] However, there are several physically important theories that do not have hyperregular, or even regular, models; most notably, gauge theories are such that the Legendre transformation defines a submanifold of T^*Q. It would be surprising, and significant, if the class of Lagrangian gauge theories and the class of Hamiltonian gauge theories were not equivalent. Therefore, it is worthwhile to consider whether one could set up an equivalence result as strong as categorical equivalence in the context of gauge theories. But to do this, we first need to define the models of the corresponding Lagrangian and Hamiltonian gauge theories. So let us start by considering the way that gauge theories are usually formulated.

2.3 The Irregular Case

We say that the Lagrangian is *irregular* when the Hessian $W_{ij} = \frac{\partial L}{\partial \dot{q}^i \dot{q}^j}$ is not invertible i.e. when it is singular. A class of irregular Lagrangian theories can be characterized by the fact that the Legendre transformation $FL(T_*Q)$ is a submanifold of T^*Q called the *primary constraint surface* Σ_p, defined by the satisfaction of a collection of (primary) constraints $\phi_a(q_i, p_i) = 0$. It is this class of irregular Lagrangian theories that we will take to constitute the gauge theories.

[5]Indeed, in a footnote, Barrett (2019) says: "One can, of course, consider the more general case, but I conjecture that there the theories will be inequivalent according to any reasonable standard of equivalence."

Given that the Legendre transformation defines a submanifold of the cotangent space in the context of gauge theories, it seems natural that we should formulate the Hamiltonian theory on this submanifold if we want to relate the two theories. Indeed, if we start with a Hamiltonian theory on T^*Q, then one can specify the theory on the primary constraint surface. First, we can define an induced presymplectic two-form $\tilde{\omega} = i^*\omega$ where $i : \Sigma_p \to T^*Q$ is the inclusion map. The null vector fields of $\tilde{\omega}$ are the vector fields corresponding to the *primary first-class constraints*, which geometrically correspond to the primary constraints whose vector field is tangent to the constraint surface.

Using this presymplectic two-form, the equations of motion on this submanifold can be written as $\tilde{\omega}(X_H, \cdot) = dH$ where H is the Hamiltonian on T^*Q restricted to the constraint surface. Notice that since $\tilde{\omega}$ is degenerate, the solutions to this equation of motion are not unique; we can think of this fact as related to the gauge nature of the theory.

This provides a well-defined theory on the primary constraint surface. However, as Dirac (1964) and others noticed, there are inconsistencies that might arise with this theory. In particular, it may not be that the primary constraints hold at all points along a solution, which corresponds to the fact that the vector fields X_H that define the solutions to this equation may not be tangent to the constraint surface. In order for the solutions to be tangent to the constraint surface, it must be that $\tilde{\omega}(X_H, Z) = dH(Z) = 0$ for vector fields Z associated with the primary constraints. But this may define a further collection of constraints that we call *secondary constraints*, and we can think of these additional constraints as leading to the specification of a further submanifold.

Continuing this process of requiring that the solutions to the equations of motion are tangent to the constraint surface terminates in a final constraint surface, $(\Sigma_f, \tilde{\omega}_f, H_f)$, defined by the satisfaction of a collection of constraints, where the null vector fields of $\tilde{\omega}_f$ are those M vector fields associated with the M first-class constraints. The integral curves of the null vector fields are called the *gauge orbits*. They are M-dimensional surfaces on the constraint surface spanned by the null vectors. In this way, on the final constraint surface, the gauge transformations are given by transformations along the integral curves of the null vector fields associated with first-class constraints.

The equations of motion $\tilde{\omega}(X_H, \cdot) = dH$ only defines X_H up to arbitrary combinations of null vectors when $\tilde{\omega}$ is presymplectic. So following standard usage, let us define the 'Total Hamiltonian' as the equivalence class of Hamiltonians defined up to arbitrary combinations of *primary* first-class constraints i.e. the equivalence class of Hamiltonians on the primary constraint surface. Similarly, we define the 'Extended Hamiltonian' as the equivalence class of Hamiltonians defined up to arbitrary combinations of *primary and secondary* first-class constraints i.e. the equivalence class of Hamiltonians on the final constraint surface.

Going forward, we will use the term 'Total Hamiltonian formalism' to refer to the formulation of irregular Hamiltonian mechanics on the primary constraint surface and 'Extended Hamiltonian formalism' to refer to the formulation of irregular Hamiltonian mechanics on the final constraint surface.

2.4 Inequivalence Argument

In the previous section, we showed that a Hamiltonian gauge theory is naturally formulated on the final constraint surface with the Extended Hamiltonian as the equivalence class of Hamiltonians. However, we also pointed out that if we start with a Lagrangian theory, the Legendre transformation defines the primary constraint surface, which corresponds to the Total Hamiltonian being the right equivalence class of Hamiltonians. This fact has led some authors to conclude that Extended Hamiltonian formalism is inequivalent to the Lagrangian formalism, and that this is reason to think that the Extended Hamiltonian formalism is mistaken.

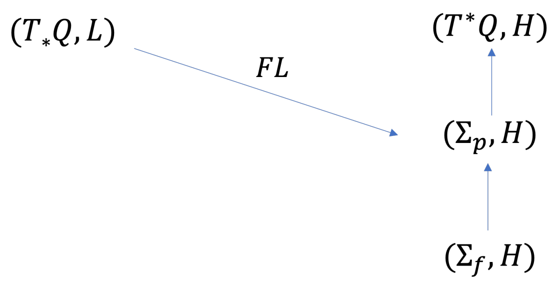

Figure 2.1: The irregular case.

For example, Gracia and Pons (1988) state that:

> "No "extended hamiltonian" is needed, since it would introduce new solutions of the equations of motion that would break the equivalence between Lagrangian and Hamiltonian formalisms".

Similarly, Pitts (2014b) argues that

> "The extended Hamiltonian breaks Hamiltonian-Lagrangian equivalence. Requiring Hamiltonian-Lagrangian equivalence fixes the supposed ambiguity permitting the extended Hamiltonian".

Such claims have been used to argue that the right definition of a gauge transformation in the Hamiltonian formalism is not given by the transformation relating solutions to the Extended Hamiltonian, but rather it is the transformation relating solutions to the Total Hamiltonian. And one can show that the transformations relating solutions to the Total Hamiltonian are not given by arbitrary combinations of first-class constraints but rather by a *particular* combination of first-class constraints, contrary to the standard definition.[6] Therefore, the claim that the Lagrangian formalism is equivalent only to the Total Hamiltonian formalism has significant implications not only for how one formulates Hamiltonian gauge theories but also for the characterization of the gauge transformations themselves.

[6]For more discussion on this debate, see Pitts (2014b); Pons (2005); Pooley and Wallace (2022).

However, to evaluate these claims, we ought to understand the sense of (in)equivalence that is at stake. This hasn't been discussed in detail in the literature; indeed, what one finds are references to certain results that are taken to show that the solutions to the Euler-Lagrange equations are equivalent to the solutions to the Hamilton-Dirac equations on the primary constraint surface. One particular result that is widely cited is found in Batlle et al. (1986), so let us spell out this result and consider the notion of equivalence that it supports.

Theorem (Batlle et al. (1986)): If $(q_i(t), \dot{q}_i(t))$ satisfies the Euler-Lagrange equations, then $FL(q_i(t), \dot{q}_i(t))$ satisfies the Hamilton-Dirac equations on the primary constraint surface. Similarly, if $(q_i(t), p_i(t))$ satisfies the Hamilton-Dirac equations on the primary constraint surface, then $FL^{-1}(q_i(t), p_i(t))$ satisfies the Euler-Lagrange equations, where $FL^{-1}(q_i(t), p_i(t))$ is constructed via:

$$\dot{q}^i = \frac{\partial H}{\partial p_i} + v_a(q_i, \dot{q}_i)\frac{\partial \phi_a}{\partial p_i}$$

$$-\frac{\partial L}{\partial q^i} = \frac{\partial H}{\partial q^i} + v_a(q_i, \dot{q}_i)\frac{\partial \phi_a}{\partial q^i}$$

where ϕ_a are the primary constraints and $v_a(q_i, \dot{q}_i)$ is arbitrary.

The theorem shows that the solutions to the Euler-Lagrange equations map to the solutions to the Hamilton-Dirac equations on the primary constraint surface and vice versa. But notice that the inverse Legendre transformation maps one point on the primary constraint surface to multiple points on the tangent space since it is defined in terms of arbitrary functions v_a. It therefore maps one solution on the primary constraint surface to multiple solutions on tangent space. If these solutions are not considered equivalent from the perspective of the Lagrangian formalism, then this result cannot establish that a Lagrangian gauge theory defined on tangent space and its corresponding Hamiltonian theory defined on the primary constraint surface have equivalent solutions.

Moreover, even if we do interpret these points/solutions as equivalent, it seems that the most that this theorem can tell us is that there is a *dynamical* equivalence between Lagrangian mechanics and Hamiltonian mechanics on the primary constraint surface. One cannot use Barrett's result to establish categorical equivalence since we do not have a way of translating the models and symmetries of one theory to those of the other. In particular, it was important for Barrett's result that $FL^{-1} = FH$, which followed from these maps being global diffeomorphisms. The maps between tangent space and the primary constraint surface do not satisfy this property. Therefore, the results in Batlle et al. (1986) are not sufficient to infer theoretical equivalence between Lagrangian gauge theories and Hamiltonian gauge theories defined on the primary constraint surface.

However, this theorem does provide the tools to infer that there is a dynamical, and therefore theoretical, *inequivalence* between Lagrangian gauge theories and the Extended Hamiltonian formalism: what the theorem shows is that the equivalence class of solutions to the Euler-Lagrange equations on tangent space are in one-to-one corresponds to the equivalence class of solutions to Hamilton's equations on the primary constraint surface. That is, once we take into account the *symmetries* of the equations of motion, then the two formalisms agree about which solutions are distinct from one another. On the other hand, the symmetries of Hamilton's equations on the final constraint surface, i.e. using the Extended Hamiltonian, are wider than symmetries of Hamilton's equations on the primary constraint surface (there are distinct solutions on the primary constraint surface that are equivalent on the final constraint surface). Therefore, the Lagrangian formalism and the Extended Hamiltonian formalism are inequivalent because their equivalence classes of solutions are different.

Indeed, it is this dynamical inequivalence that seems to be the core of the arguments that the Extended Hamiltonian gets the gauge transformations wrong, from the perspective of the Lagrangian formalism: there is a mismatch of the symmetries of the equations of motion. However, there are some lingering puzzles.

First, there is a sense in which the Total Hamiltonian formalism is *empirically equivalent* to the Extended Hamiltonian formalism: if we take secondary constraints to be a physical requirement in the Total Hamiltonian formalism, then the solutions one gets when one takes the solutions to the Total Hamiltonian and restricts to the final constraint surface are just the solutions to the Extended Hamiltonian on the final constraint surface.[7] Therefore, the fact that the equivalence classes of solutions are different doesn't seem to allow for the inference that the Extended Hamiltonian formalism is *wrong*, without some further reason to think that the Lagrangian equivalence class of solutions is the right one. Another way to put this worry is that without an account of theoretical equivalence, one cannot fully evaluate the claim that the Total Hamiltonian formalism is the *right* formulation from the perspective of the Lagrangian formalism.

Second, given that we have motivated two formulations of Hamiltonian mechanics in the presence of gauge symmetry – the Total Hamiltonian formalism and the Extended Hamiltonian formalism – it is natural to ask whether, in the context of gauge theories, one could also motivate a new formulation of *Lagrangian* mechanics whose equivalence class of solutions matches the Extended Hamiltonian formalism. If we could, then this would suggest that the dynamical inequivalence that we find between Lagrangian mechanics and the Extended Hamiltonian formalism is an accident of the way we set up the Lagrangian formalism in the first place.

These puzzles lead to the following questions: First, is there some empirically equivalent formulation of Lagrangian mechanics that is dynamically equivalent to the Extended Hamiltonian formalism? Second, can one provide a stronger account of theoretical equivalence between formulations of Lagrangian and Hamiltonian gauge theories?

[7] For a more detailed argument of this kind, see Pooley and Wallace (2022).

In what follows, I will argue that the answer to both questions is yes, and that this refutes the claim that from the perspective of (equivalence with) the Lagrangian formalism, the Total Hamiltonian formalism is motivated over the Extended Hamiltonian formalism: we can both reformulate Lagrangian mechanics such that the resulting theory is dynamically equivalent to the Extended Hamiltonian formalism, and one can set up a categorical equivalence between precisely these formulations of Lagrangian and Hamiltonian gauge theories.

More carefully, I will first demonstrate, drawing from Gotay and Nester (1979), that one can formulate Lagrangian mechanics as a constraint theory such that its models are related to models of the final Hamiltonian constraint surface in just the same way that the usual Lagrangian models are related to models of the primary Hamiltonian constraint surface. This will suffice to show that the equivalence class of solutions of this reformulated Lagrangian theory match the equivalence class of solutions of the Extended Hamiltonian formalism. I will then argue that there is a way to redefine the models of these theories using a process known as *reduction* such that one can set up a categorical equivalence result between classes of models of the reduced theories. This will demonstrate that there is a sense in which Lagrangian and Hamiltonian gauge theories are theoretically equivalent, but that this doesn't support the view that the Extended Hamiltonian formalism is wrong; to the contrary, it demonstrates that there is a natural formulation of Lagrangian mechanics that is theoretically equivalent to the Extended Hamiltonian formalism.

2.5 Lagrangian Constraint Formalism

To see how we can think of constraints in the Lagrangian formalism, let us start by writing the Euler-Lagrange equations as:

$$W_{ij}\ddot{q}^j + K_i = 0 \tag{2.1}$$

where $W_{ij} = \frac{\partial^2 L}{\partial \dot{q}^i \partial \dot{q}^j}$ is the Hessian and $K_i = \frac{\partial^2 L}{\partial \dot{q}^i \partial q^j}\dot{q}^j - \frac{\partial L}{\partial q^i}$. The singular case is characterized by the vanishing of the determinant of W_{ij}. Let us say that the rank of W_{ij} is $n - m_1$ so that W_{ij} has m_1 null vectors, φ_μ, such that $W_{ij}\varphi^j_\mu = 0$. We call these "gauge identities" because they hold at all points in T_*Q.

Contracting the equations of motion with the null vectors, we get:

$$\chi^{(1)}_\mu = K_i \varphi^i_\mu = 0 \tag{2.2}$$

We call these the first m_1 "Lagrangian constraints". We now require for consistency that these constraints are preserved under time evolution i.e. $\frac{d}{dt}\chi^{(1)}_\mu = 0$. This gives rise to new Lagrangian constraints $\chi^{(2)}_{\mu'}$. We can continue this process until we are left with all of the Lagrangian constraints. As in the Hamiltonian case, there are certain constraints whose time evolution allows one to determine some of the undetermined accelerations; as we will see, these constraints correspond to the second-class constraints on the Hamiltonian side.

It will be helpful to consider the picture more geometrically.[8] We can define, as in the regular case, the Lagrangian state space to be endowed with a two form $\Omega = FL^*\omega$ that is given in coordinate form by $\Omega = \frac{\partial^2 L}{\partial \dot{q}^i \partial q^j} dq^i \wedge dq^j + \frac{\partial^2 L}{\partial \dot{q}^i \partial \dot{q}^j} dq^i \wedge d\dot{q}^j$. When the Hessian $W_{ij} = \frac{\partial^2 L}{\partial \dot{q}^i \partial \dot{q}^j}$ is non-invertible, Ω is degenerate and so it is a pre-symplectic two-form. We call this the *irregular* case.

[8]For details, see Gotay and Nester (1979).

The geometric equations of motion can be written as before as:

$$\Omega(X_E, \cdot) = dE \tag{2.3}$$

Because Ω is not symplectic in the irregular case, there will not be a unique solution to the equations of motion; indeed there may not be any solution at some points. However, the null vectors of Ω allow us to define a submanifold where one can solve the equations at every point, in the following way. The null vectors Z of Ω are such that $\Omega(Z, \cdot) = 0$. So, in order for the equations of motion to hold, and be tangent to T_*Q, we must have that $dE(Z) = 0$. This motivates restricting to the submanifold P_1 defined by $dE(Z) = 0$ for null vectors Z. We can therefore think of $dE(Z)$ as *constraints*.

Now we require that the solutions to the equations of motion everywhere lie tangent to P_1 i.e. that the constraints hold at all points along a solution. But this is just to require that $dE(Y) = 0$ where Y is in the kernel of Ω restricted to P_1, which we can write as Ω_1. So we should restrict to a submanifold where in addition $dE(Y) = 0$. Therefore, we can think of $dE(Y)$ as further constraints.

Reiterating this process, we find a constraint surface P_k for k constraints where the solutions of the equations of motion $\Omega_k(X_E, \cdot) = dE$ are tangent to the constraint surface.[9] The null vector fields of Ω_k correspond to the null vector fields of Ω and the vector fields associated with the constraints. Therefore, we can think of this formalism as providing a way on the Lagrangian side to associate constraints with gauge transformations: the vector fields associated with the constraints generate (a subset of) the gauge transformations, understood as transformations along the integral curves of the null vector fields.

[9]Here, the energy function E should be thought of as the energy function on T_*Q restricted to the points of the constraint surface P_k.

However, there are some constraints $K_i\varphi^i_\mu = 0$ that are not accounted for by this geometric procedure. These are the constraints that do not correspond to null vector fields of the (induced) presymplectic two-forms. As Gotay and Nester (1980) show, these constraints are determined by requiring that the equation of motion is second-order, which corresponds to requiring that a solution to the equation of motion, written in coordinate-dependent form as $X = \alpha^i \frac{\partial}{\partial q^i} + \beta^i \frac{\partial}{\partial \dot{q}^i}$, is such that $\alpha^i = \dot{q}^i$ (this follows from the two-form written in coordinate form above). If constraints of this kind arise, we can find their time derivative and thereby determine potentially new constraints. So take the final constraint surface to be given by P_{k+l} with l being the number of constraints arising from the second-order condition.

2.6 Relationship between Final Constraint Surfaces

We have seen that we can construct submanifolds of the tangent bundle in a similar way to the construction of submanifolds in the Hamiltonian formalism through constraints, and that we can write the equations of motion intrinsically on these submanifolds. So the natural question is whether the theory defined on the final constraint submanifold on the Lagrangian side is equivalent to the theory defined on the final Hamiltonian constraint manifold. To present an equivalence result of this kind, we will start by using the results in Gotay and Nester (1979) to show that the relationship between the models on the final constraint manifolds is the same as the relationship one finds between the original Lagrangian model and the model on the primary constraint surface.[10]

We will restrict ourselves, following Gotay and Nester (1979), to *almost regular* Lagrangian models. An almost regular Lagrangian model is associated with two assumptions. First, FL is a submersion onto its image i.e. its differential is surjective. Second, the fibers

[10] Although the results in this section can be found in Gotay and Nester (1979), they do not discuss in detail the kind of equivalence that these results imply, nor do they draw the implications that we do here for the debate about the Total vs. Extended Hamiltonian.

$FL^{-1}(FL(q, \dot{q}))$ are connected submanifolds of T_*Q. These two assumptions guarantee that $FL^*H = E$ defines a single-valued Hamiltonian, since they imply that the energy function E is constant along the fibers $FL^{-1}(FL(q, \dot{q}))$.[11] We can think of the almost regular Lagrangian models as characterizing the Lagrangian gauge theories: they are the models of Lagrangian mechanics for which there is a well-defined corresponding Hamiltonian theory on the primary constraint surface with the Hamiltonian related to the energy function via $FL^*H = E$.

We also assume that we have no ineffective constraints[12], which means that there is a clear separation between first-class and second-class constraints i.e. a first-class constraint does not become second-class when considering its time derivative and vice versa. To start, we will assume that we just have first-class constraints on the Hamiltonian side and constraints that correspond to null vector fields on the Lagrangian side.

Let us first consider the relationship between T_*Q and the primary Hamiltonian surface Σ_p. Take i_p to be the inclusion map $i_p : \Sigma_p \to T^*Q$. Then we can define the Legendre transformation between T_*Q and Σ_p, $FL_p : T_*Q \to \Sigma_p$ via $i_p \circ FL_p = FL$. Since FL is assumed to be a submersion onto its image and its image is precisely Σ_p, FL_p is also a submersion and is surjective (but not injective nor an immersion). Moreover, take FL_{p*} is the pushforward map associated with Fl_p. The kernel of FL_{p*} (or $Ker(FL_{p*})$) is the collection of vector fields Z on T_*Q such that $FL_{p*}(Z)$ is the zero vector at all points $x \in \Sigma_p$.

> **Proposition 2.1:** Every distinct null vector field on Σ_p corresponds to a distinct null vector field on T_*Q and there are additional null vector fields on T_*Q corresponding to the vector fields in the kernel of FL_{p*}.[13]

[11] Proof can be found in Gotay and Nester (1979).

[12] An ineffective constraint is one whose gradient vanishes weakly. For discussion, see Gotay and Nester (1984).

[13] See B.1 for proof.

Proposition 2.1 tells us for every null vector field on tangent space, there is a corresponding null vector field on the primary Hamiltonian constraint surface (and vice versa), but that the relationship is many to one, with the difference in dimension of null vector fields being given by the dimension of the kernel of FL_{p*}, which is equal to the number of primary first-class constraints.

It turns out that the same relationship holds between the final constraint surfaces P_f and Σ_f. Define the induced Legendre transformation between these spaces as follows. Define $i_L : P_f \to T_*Q$ as the inclusion map from the final Lagrangian constraint surface to the tangent space and $i_H : \Sigma_f \to T^*Q$ as the inclusion map from the final Hamiltonian constraint surface to the cotangent space. Then $FL_f : P_f \to \Sigma_f$ is given implicitly by $i_H \circ FL_f = FL \circ i_L$.

> **Proposition 2.2:** Every distinct null vector field on Σ_f corresponds to a distinct null vector field on P_f and there are additional null vector fields on P_f corresponding to the vector fields in the kernel of FL_{f*}.[14]

Proposition 2.2 tells us that the relationship between null vector fields on the final constraint surfaces is such that the number of null vector fields on P_f is equal to the number of null vector fields on Σ_f plus the dimension of $Ker(FL_{f*})$. One can also show that $Ker(FL_{f*}) = Ker(FL_{p*})$, and so $Ker(FL_{f*})$ has dimension equal to the number of primary first-class constraints.

Finally, we can show that the solutions to the equations of motion are related in a similar way, using the fact that $FL_f^*(H) = E$ on the final constraint surfaces:

> **Proposition 2.3:** Every distinct solution to $\tilde{\omega}_f(X_H, \cdot) = dH$ on Σ_f corresponds to a distinct solution to $\Omega_f(X_E, \cdot) = dE$ on P_f and there are additional solutions to $\Omega_f(X_E, \cdot) = dE$ related by vector fields in the kernel of FL_{f*}[15]

[14] See B.2 for proof.
[15] See B.3 for proof.

Proposition 2.3 implies that every solution to the Lagrangian equations of motion on the final constraint surface corresponds to a solution to the Hamiltonian equations of motion on the final constraint surface and vice versa. Moreover, there is not a one-to-one correspondence of solutions in the same way that there is no one-to-one correspondence of null vector fields.

This shows that the relationship between the Lagrangian and Hamiltonian theories defined on the final constraint surfaces is the same as the relationship between the theory defined on T_*Q and the theory defined on the primary constraint surface: we can map solutions to solutions, but only up to symmetries on the Lagrangian side, where the symmetries are generated by null vector fields. Therefore, we can say that the theories formulated on the final constraint surfaces are dynamically equivalent, in the sense that they agree on the equivalence class of solutions. This provides a partial response to the claim that the Extended Hamiltonian formalism is inequivalent to the Lagrangian formalism: there is in fact an alternative formulation of Lagrangian gauge theories that is dyanmically equivalent to the Extended Hamiltonian formalism in the same way that the original formulation of Lagrangian mechanics is dynamically equivalent to the Total Hamiltonian formalism.

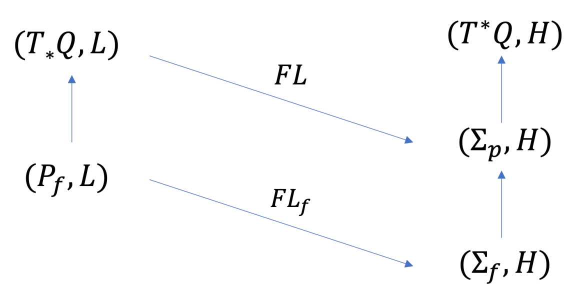

Figure 2.2: Relationship between final constraint surfaces.

However, we do not yet have a way to provide a stronger, theoretical equivalence result. In particular, we need a way of characterizing the claim that X_E is equivalent to $X_E + Y_i$ for $Y_i \in Ker(FL_{f*})$, or more generally, we need a way of characterizing the structure of the theories that includes the transformations generated by the null vectors. Moreover, although we have defined a map from the final Lagrangian constraint surface to the final Hamiltonian constraint surface, and we can use this to pull back structures from the Hamiltonian model to the Lagrangian model, we do not yet have the tools to set up a categorical equivalence result analogous to the result in Barrett (2019) since FL_f is not a diffeomorphism and so it is not related to the fiber derivative of the Hamiltonian on the constraint surface in the right way.

Before turning to how we might set up a categorical equivalence result, let us consider how the situation changes when we also have second-class constraints on the Hamiltonian side. Since we assumed that there are no ineffective constraints, this means that we only need to consider the case where we have primary second-class constraints, since the time derivative of these constraints will generate any additional second-class constraints.

We have shown that we can relate the first-class constraints to null vector fields on the Lagrangian side. But since second-class constraints do not correspond to null vector fields, we cannot relate them to a Lagrangian constraint in the same way. However, it turns out that for every (distinct) primary second-class Hamiltonian constraint, there is a corresponding (distinct) Lagrangian constraint whose associated vector field is not null. In particular, the additional Lagrangian constraints are the pullback under the (induced) Legendre transformation of the time derivative of a second-class Hamiltonian constraint.[16] Generalizing, the final Lagrangian constraint surface will be reduced by the number of second-class constraints on the Hamiltonian side.

2.7 Reduction and Equivalence

Although we now have a picture under which both the Lagrangian formalism and the Hamiltonian formalism can be written intrinsically on constraint manifolds that are systemically related, we do not yet have a theoretical equivalence result. Recall that the barrier is that

[16]For details, see Batlle et al. (1986); Pons (1988).

we do not have a way to define a translation from Lagrangian to Hamiltonian models and vice versa via the relationship between FL, FL^{-1}, FH and FH^{-1} since the final constraint submanifolds are not of the same dimension.

However, there is an indication that we should be able to set up an equivalence result: while the dimensions of the final constraint surfaces are different, the difference seems to be due to *arbitrariness* in the Lagrangian formalism coming from the null vector fields in the kernel of FL_{f*}. Indeed, if we take null vector fields to generate symmetries, then there is an argument that once we have accounted for all of the symmetries, the two formalisms are in complete agreement. One way of thinking about 'accounting for the symmetries' is to consider whether there is a way to characterize the theories in terms of the equivalence class of states along the integral curves of the null vector fields. In fact, there is a well-known construction for specifying a Hamiltonian theory in terms of the equivalence class of states called *reduction*: the process of reduction defines a manifold that "quotients out" the gauge transformations. This is not a construction that one often finds discussed for a Lagrangian theory.[17] However, we have shown that we can think of a Lagrangian gauge theory in an analogous way to the Hamiltonian formalism as defined on a pre-symplectic manifold. This suggests that we should be able to equally construct a reduced space for the final Lagrangian constraint surface. The question then becomes: are the *reduced* versions of Lagrangian and Hamiltonian gauge theories categorically equivalent?

The reason that reduction will help us to set up a categorical equivalence result is that one can show that reduction induces a symplectic two-form on the reduced space. Recall that being symplectic means that the Lagrangian/Hamiltonian models are *regular*: the two-form is non-degenerate and so we can, at least locally, define the inverse of the fiber derivatives. Therefore, if we can show that the Legendre transformation of a reduced Lagrangian model gives rise to a reduced Hamiltonian model and vice versa, then this suggests that we can set

[17] An exception is Pons et al. (1999).

up an equivalence result in an exactly analogous way to Barrett (2019), if we restrict to the hyperregular reduced models.

In order to show that this is indeed possible, we will show that the dimensions of the reduced spaces related by FL_f are the same, that the structures defined on this space can be inherited from the final constraint surface in a natural way, and that the image of the Legendre transformation of the reduced Lagrangian space is precisely the corresponding reduced Hamiltonian space. These will provide the tools to prove categorical equivalence between classes of models of the reduced theories.

Consider first a presymplectic Hamiltonian manifold $(\Sigma, \tilde{\omega}, H)$ that is foliated by the gauge orbits at each point. We can define a smooth, differentiable manifold $\bar{\Sigma}$ by taking the quotient of Σ by the kernel of $\tilde{\omega}$ i.e. the null vector fields of $\tilde{\omega}$. Recall that the integral curves of the null vector fields define the gauge orbits, and so the points of the quotient manifold are just the equivalence class of points along the gauge orbits. This is well-defined since the gauge orbits foliate the constraint surface in such a way that one can define a transverse manifold that meets each leaf of the foliation in at most one point i.e. through each point there is only one gauge orbit.[18] Recall that on the final constraint surface, the dimension of the gauge orbits is the number of first-class constraints M_f and the dimension of Σ_f is $2N - M_s - M_f$ where N is the dimension of configuration space and M_s is the number of second-class constraints. So the quotient manifold of the final Hamiltonian constraint surface $\bar{\Sigma}$ has dimension $2N - M_s - 2M_f$.

Define an open, surjective projection map $\pi : \Sigma_f \to \bar{\Sigma}$ such that we define the reduced two-form $\bar{\omega}$ via $\tilde{\omega}_f = \pi^*(\bar{\omega})$, which acts according to $\bar{\omega}(\bar{X}, \bar{Y}) = \tilde{\omega}_f(X, Y)$. One can show that $\bar{\omega}$ is well-defined and is symplectic.[19] We can also define a reduced Hamiltonian $\bar{H}$ as

[18] See Souriau (1997) §5 ans §9 for details.

[19] It is well-defined since the value of $\tilde{\omega}_f$ doesn't depend on which point along the gauge orbit one considers. It is closed since $\tilde{\omega}_f$ is closed and π is a surjective submersion, and it is non-degenerate since $Ker(\bar{\omega}) = Ker(\tilde{\omega}_f)/Ker(\tilde{\omega}_f) = 0$.

the value of H on the equivalence class of points along the gauge orbits i.e. $H = \pi^*(\bar{H})$. This is well-defined because H is constant along the gauge orbits on the final constraint surface (since the solutions to the equations of motion are tangent to the final constraint surface). We can therefore write the equations of motion on the reduced space in terms of the reduced Hamiltonian $\bar{H}$ as $\bar{\omega}(X_{\bar{H}}, \cdot) = d\bar{H}$, and the solutions are just the projection of the solutions to the equations of motion on Σ_f to $\bar{\Sigma}$: they are just the solutions defined for the gauge-invariant quantities.

To summarize, there is a well-defined Hamiltonian theory on the reduced space of the final Hamiltonian constraint surface in terms of a symplectic two-form and a reduced Hamiltonian function. However, this only required that we had a presymplectic manifold with a foliation induced by the null vector fields of the associated two-form and that the Hamiltonian function was constant along the gauge orbits. Given that the same is true for the Lagrangian final constraint surface, we can do the same reduction procedure on the Lagrangian side to produce a reduced Lagrangian space. This will have dimension $2N - 2L_a - L_b$ where L_a is the number of Lagrangian constraints associated with null vector fields and L_b is the number of additional Lagrangian constraints. As in the Hamiltonian case, because the energy function E is constant along gauge orbits on the final constraint surface, the reduced Lagrangian function $\bar{L}$ will be well-defined as well. We can therefore write the equations of motion as $\bar{\Omega}(X_{\bar{E}}, \cdot) = d\bar{E}$ where $\bar{E}$ is the energy function associated with $\bar{L}$ and $\bar{\Omega}$ is the reduced symplectic two-form.

Let us now turn to the relationship between the models of the reduced theory. First, let us consider the relationship between the dimensions of the reduced spaces corresponding to models on the final constraint surfaces P_f, Σ_f that are related via FL_f. Recall that the dimension of the Lagrangian final constraint surface P_f is equal to the dimension of the Hamiltonian final constraint surface Σ_f plus the number of primary first-class constraints. But recall also that the dimension of the kernel of Ω_f is equal to the number of first-class constraints plus the number of primary first-class constraints. Therefore, the dimension of the reduced Lagrangian space $\bar{P}$ is equal to the dimension of the Hamiltonian constraint surface Σ_f minus the number of first-class constraints. But this is just the dimension of the reduced Hamiltonian space, $\bar{\Sigma}$. Therefore, the dimensions of the reduced Lagrangian final constraint surface and the reduced Hamiltonian final constraint surface are equal.

Now define an induced transformation $\bar{F} : \bar{P} \to \bar{\Sigma}$ that satisfies $\pi_H \circ FL_f = \bar{F} \circ \pi_L$ where $\pi_H : \Sigma_f \to \bar{\Sigma}$ and $\pi_L : P_f \to \bar{P}$ are the projection maps. This provides a way to map from the reduced Lagrangian space to the corresponding reduced Hamiltonian space. Moreover, notice that since $\bar{L}$ is regular (since the induced two-form is symplectic), the Legendre transformation on $\bar{P}$ will be a local diffeomorphism. And since $\bar{P}$ and $\bar{\Sigma}$ have the same dimension, the induced transformation $\bar{F}$ is precisely the Legendre transformation on $\bar{P}$, $F\bar{L}$. That is, $\bar{F} : \bar{P} \to \bar{\Sigma}$ is the Legendre transform on T_*Q, FL, projected to the reduced space. Similarly, since $\bar{H}$ is regular, the fiber derivative of $\bar{H}$, $F\bar{H}$, will be a local diffeomorphism and it will map $\bar{\Sigma}$ to $\bar{P}$. Using the reduced Legendre transformation, one can also show that the reduced symplectic two-forms are related via $F\bar{L}^*(\bar{\omega}) = \bar{\Omega}$ and the reduced Hamiltonian and energy function are related via $F\bar{L}^*(\bar{H}) = \bar{E}$.[20]

Finally, since (P_f, L_f) is, by assumption, an almost regular system, $(\bar{P}, \bar{L})$ will also be almost regular. This implies that $F\bar{L}$ is injective [21]. Moreover, the image of $F\bar{L}$ is $\bar{\Sigma}$ by construction so $F\bar{L}$ is surjective. But this means that $F\bar{L}$ is a global diffeomorphism, and so $(\bar{P}, \bar{L})$ is in fact a *hyperregular* system. Therefore, we can define the inverse $F\bar{L}^{-1} : \bar{\Sigma} \to \bar{P}$. This allows us to define $\bar{H} = \bar{E} \circ F\bar{L}^{-1}$.

Therefore, for an almost regular Lagrangian model defined on the final constraint surface, we can construct a reduced model such that this model is hyperregular and its Legendre trans-

[20] To see this, notice that $\pi_L^*(\bar{FL}^*\bar{\omega}) = FL_f^*(\pi_H^*\bar{\omega}) = FL_f^*\tilde{\omega}_f = \Omega_f$. Since π_L is a surjective submersion, this implies that $F\bar{L}^*(\bar{\omega}) = \bar{\Omega}$. The second follows by similar reasoning.

[21] The reason is that for an almost regular system the image of the Legendre transformation is the leaf space of the foliation generated by the kernel of the pushforward of the Legendre transformaiton. When a system is regular, this kernel is zero, and so it must be injective.

formation is precisely the (hyperregular) reduced model of the corresponding Hamiltonian final constraint surface. This implies that as long as we are concerned with almost regular Lagrangian models and their corresponding Hamiltonian models, the reduced formulations of these theories bear exactly the same relationship as hyperregular models of Lagrangian and Hamiltonian mechanics.

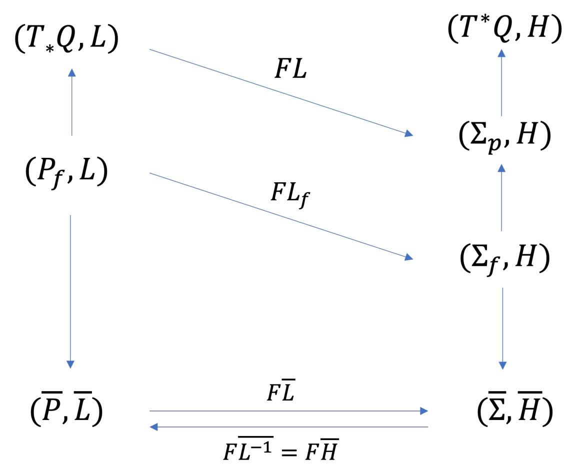

Figure 2.3: Relationship between reduced spaces.

We are now at the point where we can set up an equivalence result. Recall that in order to do so, we need to define the models and symmetries of the associated theories. In the hyperregular case given by Barrett (2019), the symmetries were the point-transformations that preserved the Lagrangian/Hamiltonian. However, in order for the point-transformations to be well-defined for the reduced theories, we need that the reduced state space has the form of a (co)tangent bundle. This is not guaranteed by the above; at least, it will depend upon the particular nature of the constraints and the gauge transformations.[22] On the other hand, we do have that the reduced spaces are symplectic manifolds. Therefore, it seems that the natural notion of symmetry is rather given by *symplectomorphisms:* diffeo-

[22]Moreover, even if one could think of the reduced state space as having the structure of (co)tangent space, it isn't clear that one would want the symmetries to be given by point-transformations. As Barrett (2015b) shows, there are point$_*$-transformations that don't preserve an arbitrary symplectic two-form on T_*Q. One might conclude from this that point$_*$-transformations are not the relevant symmetries to consider for the reduced Lagrangian models, since the symplectic two-form is an integral part of the construction of these reduced models.

morphisms that preserve the symplectic two-form on the reduced space (and preserve the Lagrangian/Hamiltonian).

So let us define the category **LagR** as having objects $(\bar{P}, \bar{\Omega}, \bar{L})$ and take the arrows between objects $(\bar{P}_1, \bar{\Omega}_1, \bar{L}_1)$ and $(\bar{P}_2, \bar{\Omega}_2, \bar{L}_2)$ to be given by symplectomorphisms i.e. diffeomorphisms $f : \bar{P}_1 \to \bar{P}_2$ such that $f^*(\bar{\Omega}_2) = \bar{\Omega}_1$, that preserve the Lagrangian in the sense that $f^*\bar{L}_2 = \bar{L}_1$.

Similarly, define the category **HamR** as having objects $(\bar{\Sigma}, \bar{\omega}, \bar{H})$ and take the arrows between objects $(\bar{\Sigma}_1, \bar{\omega}_1, \bar{H}_1)$ and $(\bar{\Sigma}_2, \bar{\omega}_2, \bar{H}_2)$ to be given by symplectomorphisms $g : \bar{\Sigma}_1 \to \bar{\Sigma}_2$ such that $g^*(\bar{\omega}_2) = \bar{\omega}_1$, that preserve the Hamiltonian in the sense that $g^*\bar{H}_2 = \bar{H}_1$.

Define the functor J that takes the object $(\bar{P}, \bar{\Omega}, \bar{L})$ to $(\bar{\Sigma}, \bar{\Omega} \circ F\bar{L}^{-1}, \bar{E} \circ F\bar{L}^{-1})$ and that takes the arrow $f : \bar{P}_1 \to \bar{P}_2$ to $F\bar{L}_2 \circ f \circ F\bar{L}_1^{-1}$. Similarly, define the functor K that takes models $(\bar{\Sigma}, \bar{\omega}, \bar{H})$ to $(\bar{P}, \bar{\omega} \circ F\bar{H}^{-1}, (\bar{\theta}_a(X_{\bar{H}})^a - \bar{H}) \circ F\bar{H}^{-1})$ where $\bar{\theta}$ is the reduced one form, and arrows $g : \bar{\Sigma}_1 \to \bar{\Sigma}_2$ to $F\bar{H}_2 \circ g \circ F\bar{H}_1^{-1}$.

Proposition 2.4: J : **LagR** $\to$ **HamR** and K : **HamR** $\to$ **LagR** are equivalences that preserve solutions.[23]

2.8 Upshots

Proposition 2.4 tells us that there is a formulation of irregular Lagrangian mechanics that is theoretically equivalent to a formulation of irregular Hamiltonian mechanics. More precisely, it tells us that the categories of hyperregular reduced models of the final constraint surfaces are equivalent. This is significant for several reasons.

[23]See B.4 for proof.

First, we discussed in Section 2.3 the view that the correct Hamiltonian formulation is the Total Hamiltonian formalism on the basis that it is equivalent to the Lagrangian formalism in the context of gauge theories. But our arguments have suggested that in fact the Extended Hamiltonian formalism can be motivated in a similar, and even stronger, way: there are reasons to move to the final Lagrangian constraint surface from the perspective of the Lagrangian formalism, and not only can the models formulated on the Lagrangian final constraint surface be said to be dynamically equivalent to models of the Extended Hamiltonian formalism, one can also give a stronger, theoretical equivalence result between the reduced versions of such models.

In order to deny that such results provide support for the Extended Hamiltonian formalism, one would have to maintain that there is something mistaken about the Lagrangian constraint formalism that we presented. One avenue might be to argue that we shouldn't think of Lagrangian constraints as imposing a restriction on the state space of Lagrangian mechanics: they should be thought of as *dynamical* constraints and not *kinematical* constraints, and therefore they should place a restriction only on the dynamically possible models and not the kinematically possible models. In this view, the correct formulation of the kinematically possible models is given by the usual tangent bundle formulation and the formulation on the Hamiltonian primary constraint surface. In further support for this view, one might point to the fact that the categorical equivalence result that we presented goes through for this characterization of irregular Lagrangian and Hamiltonian models: it corresponds to the special case where there are no Lagrangian/secondary constraints.[24]

Although this response highlights interesting questions about the role of kinematics vs. dynamics in evaluating constraints, I think that there are good reasons to think that this distinction is not significant. For one, the dynamical solutions that we get are the same

[24]However, such an equivalence result is complicated by the fact that the solutions to the equations of motion are not tangent to the constraint surface in the case where there are Lagrangian/secondary constraints that are not represented in the structure of the state space, and so the reduced equations of motion are not well-defined. See Pons et al. (1999) for further discussion.

whether we define the equations of motion intrinsically on the final constraint surface or we consider the equation of motion on the tangent bundle and then impose the constraints. Therefore, there isn't any clear empirical difference between these formulations. Second, there is a natural sense in which the formulations on the final constraint surfaces have less structure: there are more symmetries of the theories formulated on the final constraint surfaces than on the tangent bundle/primary constraint surface since there are more null vector fields.[25] And so, if one is motivated by parsimony considerations, it is natural to think that the final constraint surface is the right intrinsic formulation of the theory.

Second, showing that there is an equivalence between Lagrangian and Hamiltonian gauge theories suggests that it is wrong to view one formulation as more fundamental than the other since they have the same underlying structure. This is interesting because the usual way of setting up the Hamiltonian formalism in the presence of constraints is by starting with a Lagrangian formulation and using it to define the primary constraints and Total Hamiltonian, which suggests that the Lagrangian formulation is more fundamental. On the other hand, the equivalence result suggests that one can instead start with a Hamiltonian theory with constraints, reduce the final constraint surface, and use this to define the corresponding Lagrangian theory.

Moreover, it is often assumed that in order to find the gauge-invariant degrees of freedom, one ought to use the Hamiltonian formulation. For example, Earman (2002b) says: "Is there then some non-question begging and systematic way to identify gauge freedom and to characterize the observables? The answer is yes, but specifying the details involves a switch from the Lagrangian to the constrained Hamiltonian formalism." The reason is that the constrained Hamiltonian formulation draws out the connection between constraints and gauge symmetry: the gauge transformations are those transformations generated by arbitrary combinations of first-class constraints, and we can define the observables as just those quantities whose

[25] cf. Section 1.8.1.

Poisson bracket with the first-class constraints is zero. In the usual way of expressing the Lagrangian formulation, we find the symmetries by using Noether's second theorem, which doesn't directly connect the idea of constraints and observables. The geometric formulation shows that if the focus is on the null vector fields of the associated two-form, then the Lagrangian formulation draws out the gauge transformations in the same way.

However, there are several subtleties with the equivalence result given by Proposition 2.4. For one, we restricted to a subset of the Lagrangian models, the 'almost regular' ones, and then considered the corresponding Hamiltonian models defined via the Legendre transformation. While we were able to show that the almost regular Lagrangian models have hyperregular reduced models, and therefore that the Hamiltonian models defined from these models also have hyperregular reduced models, we did not show that this exhausts the class of hyperregular reduced models. It would therefore be interesting to consider whether there are hyperregular reduced models that cannot be thought of as coming from a 'gauge theory' in the sense of being an almost regular Lagrangian model or its corresponding Hamiltonian model. Moreover, 'almost regularity' referred to the Lagrangian model, but there doesn't seem to be a clear Hamiltonian analogue: the fiber derivative of the Hamiltonian on the primary/final constraint surface does not construct an almost regular Lagrangian model. Therefore, it seems that we need some alternative way to characterize the relevant class of gauge theories in Hamiltonian terms.[26] These subtleties suggest that there is more work to be done in motivating the physical reasonableness of restricting to hyperregular reduced models to show equivalence between irregular models of Lagrangian and Hamiltonian mechanics.

Another subtlety of the equivalence result is that symplectic reduction can lead to counterintuitive conclusions, which has led several authors to argue that one should not reduce gauge theories (at least in certain contexts). The most notable example of this is the *Problem*

[26]For example, is it the case that any regular Hamiltonian theory with the addition of constraints give rise to a constraint surface model that is the Legendre transformation of some almost regular Lagrangian model?

of Time: when one reduces theories that are time reparameterization invariant, one ends up with a theory with no meaningful notion of evolution. If one finds these arguments convincing, then one might think that the equivalence result given by Proposition 2.4 is irrelevant; what matters is not whether the reduced theories are equivalent, but whether the unreduced theories are.

I take this to be an important limitation of the arguments presented here. However, one response is to point out that all one has done by moving to the theory formulated on the reduced space is to equivocate between states/solutions that are symmetry-related in the theory formulated on the final constraint surface. Therefore, if we interpret symmetry-related states/solutions as equivalent, then arguably the theories defined on the final constraint surface and on the reduced space have the same (symmetry-invariant) content. This suggests that even if one doesn't have a categorical equivalence result directly between classes of models on the final constraint surface, one can infer that they are equivalent from the fact that the reduced theories are equivalent.[27]

There is an interesting connection here to another strand of literature: the difference between 'reduction' and 'sophistication' (Dewar (2019)). A sophisticated version of a theory is, broadly, one where all the transformations that we take to be symmetries are isomorphisms of the models of the theory. Dewar (2019) conjectures that the sophisticated and the reduced versions of a theory are categorically equivalent. Here, we have defined and compared the reduced versions of Lagrangian and Hamiltonian gauge theories. But what do the corresponding *sophisticated* versions of the theories look like? Arguably, the theories formulated on the final constraint surfaces are 'sophisticated', in the sense that the symmetries – the gauge transformations – are isomorphisms of the models of the final constraint surface. This will be explored further in Chapter 4.

[27]Indeed, I think one *could* spell out a categorical equivalence result directly between classes of models on the final constraints surface. However, setting up such a functor is less clear than it is for the reduced theories, which is why this was not the approach taken in this chapter.

2.9 Conclusion

To conclude, I have argued that there is a sense in which Lagrangian and Hamiltonian gauge theories are equivalent by showing that one can formulate these theories geometrically on a presymplectic constraint manifold such that the hyperregular class of reduced models of these constraint models are categorically equivalent and agree dynamically. This provides an extension to the result in Barrett (2019) that hyperregular Lagrangian and Hamiltonian theories are categorically equivalent. Moreover, this extension sheds light on philosophical debates regarding the definition and interpretation of gauge transformations. In particular, in showing that one could motivate a formulation of Lagrangian gauge theories that is equivalent to Extended Hamiltonian formalism, we thereby demonstrated that the Extended Hamiltonian can be motivated from the perspective of the Lagrangian formalism, contrary to claims found in the literature.

However, this equivalence result relied on several important assumptions that are not relevant in the case considered by Barrett (2019). First, it depended on how we understand the role of constraints in the construction of the models of the theories. Second, it depended upon an interpretation of the null vector fields of a presymplectic two-form as generating the (gauge) symmetries of the theory. Finally, it depended upon reduction, and restricting to the class of hyperregular reduced models, being justified. Inasmuch as all of these assumptions are disputable, there remain interesting questions regarding the relationship between Lagrangian and Hamiltonian gauge theories.

Moreover, while categorical equivalence suggests that we can move back and forth interchangeably between Lagrangian and Hamiltonian gauge theories, there were several subtleties regarding the way that we defined the categories of the models of these theories that suggest possible avenues for maintaining that one framework is a more natural expression of gauge theories than the other. For example, the (pre)symplectic structure of Lagrangian mechanics was motivated by thinking about the Hamiltonian (pre)symplectic structure, and the class of Hamiltonian models for which categorical equivalence held were defined in terms of the Lagrangian models that they were related to. Whether one should think that (pre)symplectic is faithful to the structure of Lagrangian mechanics, and whether one can motivate the class of Lagrangian gauge theories in terms of Hamiltonian quantities, are open questions that would further deepen the understanding of the relationship between Lagrangian and Hamiltonian gauge theories.

Chapter 3

The Representational Role of Sophisticated Theories

3.1 Introduction

How should one remove "excess structure" from a physical theory? Dewar (2019) presents two ways to undertake such a task: first, one could move to a *reduced* version of the theory, where the models of the reduced theory are specified only in terms of structure that is invariant under the symmetries of the original theory. Second, one could move to a *sophisticated* version of the theory, where one defines additional maps between models of the original theory that preserve the structure invariant under the relevant symmetries such that symmetry-related models can be regarded as isomorphic. Dewar argues that despite these alternatives attributing the same structure to the world, the sophisticated version can have explanatory benefits over the reduced version.

Here, I provide a different argument in favour of sophistication: there are concrete cases where distinct physical situations are more naturally represented by the sophisticated theory than the reduced theory. The reason, I argue, is that the models of a sophisticated theory have further resources than the models of a reduced theory for representing additional detail about physical situations such that the sophisticated theory can draw more physical distinctions than the reduced theory. While it has been argued elsewhere that isomorphic models can be used to represent distinct situations,[1] these arguments do not directly show that the reduced version of a theory is representationally lacking. Indeed, if the reduced version of a theory posits the same structure as the sophisticated alternative, how can the sophisticated version represent a greater number of physical distinctions?

I will argue that this tension can be resolved by considering more carefully the ways that isomorphic models can be used to represent distinct situations. I present a division between two kinds of structure in the context of a theory that I call *theoretical structure* and *auxiliary structure*, and I demonstrate that auxiliary structure can be used to play a representational role in the sophisticated theory in a way that is absent in the reduced theory. In particular, I will argue that the sense in which isomorphic models of a sophisticated theory can represent distinct situations is that one has the freedom to define auxiliary structure within these models that can be used to represent a physical standard of comparison. It is the ability to fix this additional structure that in certain cases is lacking in the reduced version of the theory, since one may not have the same representational freedom associated with auxiliary structure within the models of the reduced theory. This will provide support for the claim that the sophisticated version of a theory can be representationally advantageous to its reduced version, without rejecting the claim that the sophisticated and reduced versions of a theory ascribe the same structure, when this structure refers to *theoretical structure*.

[1]For instance, Belot (2018); Fletcher (2020); Roberts (2020).

3.2 Removing Excess Structure

The background situation is that we have a theory that we believe has excess structure, in the following sense: there are models of the theory that are distinct, or more precisely, *non-isomorphic*, that nonetheless are related by a symmetry.[2] The symmetry between these distinct models is such that we think these models ought to be taken to represent a single physical situation. And yet, in being distinct models, the natural interpretation of the theory is that these models represent distinct physical situations. Therefore, the theory attributes structure to the world that we do not believe corresponds to anything physical.

There are several interpretive assumptions in the background of this argument for the presence of excess structure.[3] For one, there is an assumption that when models are non-isomorphic, they cannot be interpreted as representing the same physical situation. This relies on taking seriously the claims of the theory in the sense that differences between the models of the theory are taken to correspond to differences in the physical situations represented by such models. Inasmuch as we take the aim of physical theories to be that they capture the structure that the world and its parts have, this seems to be a natural assumption. Notice, however, that the argument does not seem to require that we take our theory to be a full description of all the structure of the world; only that the structure it does describe corresponds to structure in the world. We will return to this point later on.

Given that we find ourselves in such a situation, the aim is to find a way to adapt the theory such that it no longer has excess structure. In a recent paper, Dewar (2019) presents two ways to do this.

The first is *reduction*. Reduction alters the syntax of a theory by defining the theory in terms of quantities that are invariant under the relevant symmetries. This effectively equivocates

[2]Precisely how to characterize this symmetry is debatable. Broadly, the symmetry is such that it preserves the dynamics of the theory. See Belot (2013) for discussion.

[3]For further discussion, see Earman (2004); Ismael and Van Fraassen (2003); Dasgupta (2016).

between the non-isomorphic models related by the relevant symmetries in the original theory such that they correspond to identical models in the reduced theory. This is arguably the standard method for removing excess structure from a theory.

The second is *sophistication*. Sophistication keeps the syntax of the theory the same but instead alters the semantics of the theory with respect to which the theory is interpreted so that the relevant symmetries act as isomorphisms under the new semantics.[4] Inasmuch as isomorphisms preserve the structure of the models of a theory, the natural interpretation of isomorphic models is that they correspond to a single physical situation. Therefore, one has removed excess structure not by redefining the models of the theory, but by redefining the symmetries of the theory.[5]

To see this distinction in action, we will consider an example from Dewar (2019) of a theory that describes physical situations in which every object is 'handed'. Call this theory T_H in the signature $\Sigma = \{L, R\}$ and take it to be described by the following axioms, which say that every object is (excusively) either 'left' or 'right' handed:

$$\forall x(Lx \vee Rx)$$
$$\forall x \neg(Lx \wedge Rx)$$

This theory has excess structure in the following sense. Consider, for example, two models of the theory with the same domain where in one model all objects have the property L while in the second model all objects have the property R. These models are non-isomorphic

[4]One can spell out sophistication as introducing additional maps ('arrows') into the category of models of the theory and specifying their inverse and compositions with other maps. This procedure is described in Weatherall (2016a), although Dewar (2019) coins the term 'sophistication' for this procedure.

[5]There is a more fine-grained distinction that Dewar gives between 'internal' and 'external' sophistication, where the difference is whether one uses new mathematical tools to define the models of the theory such that there is a natural isomorphism between them, or whether one stipulates what counts as an isomorphism. This distinction will not be crucial for the purposes of this chapter.

within T_H; the map taking L to R and vice versa doesn't preserve the extension of these properties. And yet, these models are symmetry-related: intuitively, if we imagine the physical situation represented by one of these models, it is indistinguishable from the physical situation represented by the other model; both models correspond to a situation where everything is handed in the same way. More generally, any two models that are related by 'flipping' the handedness for each object are symmetric in this way. Therefore, there are distinct models in the theory that are symmetry related, and this indicates that T_H has excess structure.

To describe a reduced version of the theory, T_R, we need to define the theory in terms of quantities that are invariant under the relevant symmetry. An obvious candidate is a congruence relation, Cxy, that specifies whether two objects have the same handedness or not. It can be defined via:

$$\forall x \forall y (Cxy \Longleftrightarrow ((Lx \wedge Ly) \vee (Rx \wedge Ry)))$$

We can then specify T_R in terms of axioms for Cxy that say that this relation is an equivalence relation with two equivalence classes.

To define a sophisticated version of the theory, T_S, we need to alter the semantics of T_H such that the symmetry-related but non-isomorphic models of T_H are treated as isomorphic. We can do this by defining an invertible homomorphism h from a model m to a model n that consists of a map $h_1 : |m| \to |n|$ and a bijection $h_2 : \{L, R\} \to \{L, R\}$ such that:

$$h_1[L^m] = (h_2(L))^n$$

$$h_1[R^m] = (h_2(R))^n$$

This homomorphism is such that it can map 'left' hands to 'right' hands across models (and vice versa).[6] In this way, the homomorphism need not preserve the extension of the properties given by L and R and so there is no longer a well-defined notion of trans-model identity for the terms 'left' and 'right'. Thus, while models related by a change in handedness for all objects were non-isomorphic in T_H, they are isomorphic in T_S.

3.2.1 Dewar's Two Claims

There are two central claims that Dewar (2019) makes regarding the comparison between the reduced and sophisticated versions of a physical theory:

1. The reduced and sophisticated versions of a theory are theoretically equivalent.
2. The sophisticated version of a theory has explanatory benefits compared to the reduced version.

To unpack the first claim, we need a better grasp on what it is for two versions of a theory to be "theoretically equivalent". Dewar (2019) argues that the reduced and sophisticated versions of a theory are equivalent in the sense of *categorical equivalence*:

> Take the category of models of the reduced theory to be $Mod(T_R)$ and the category of sophisticated models to be $mod(T)$. Then, there is a (reasonable) functor $F : mod(T) \to Mod(T_R)$ that is full, faithful, and essentially surjective.

[6] Following footnote 5, this is the external way of defining the sophisticated theory.

Categorical equivalence as a notion of theoretical equivalence has been defended elsewhere,[7] but the important feature of categorical equivalence here is that it captures the relationship between models; in particular, the equivalence between models that are stipulated to be isomorphic in a sophisticated theory.

Turning to the second claim, Dewar (2019) says that:

> "The reduced theory treats the invariant quantities Q as primitives; this means that if some $q \in Q$ obeys some non-trivial condition as a result of its definition (in the unreduced theory), it must be asserted to obey that condition (in the reduced theory) as a simple posit." (p.496)

In other words, the reduced theory comes at an explanatory loss in the sense that the reduced theory must stipulate certain conditions that fall out naturally from the unreduced theory. Since the sophisticated version of the theory does not change the syntax of the theory, it does not come with the same loss.

In the handedness theory, for example, there is a fact in the sophisticated theory that must be assumed in reduced theory: the claim that the congruence relation Cxy is symmetric. In the reduced theory, this claim is one of the axioms, while in the sophisticated theory, it follows automatically from the axioms. Therefore, on the above account, the sophisticated theory explains this claim better.

There are some immediate worries regarding the explanatory account. First, one might question the significance of some fact being a posit in a theory rather than a consequence of a theory's axioms for the purpose of assessing the overall quality of some theory. Second, one might argue that the reduced theory also has explanatory benefits. For example, while the equivalence between certain models is stipulated in the sophisticated theory, it follows

[7] See Weatherall (2019b) for an overview and references therein.

automatically if one takes the physical structure to be that given by the reduced theory. If both the reduced and sophisticated versions of a theory have explanatory benefits, how do we weigh them up to determine which theory is "better"?[8]

In what follows, I will not aim to resolve these worries. Instead, I will provide an alternative reason to prefer sophisticated theories and suggest that it is a more robust argument in favor of sophistication by being less susceptible these worries. One might see this alternative reason as supporting a version of the explanatory account; I will discuss this possibility briefly at the end of Section 3.3.

3.3 The Representational Benefit of Sophistication

The alternative reason to prefer sophisticated theories is that they can have *representational* benefits: Isomorphic models in a sophisticated theory that correspond to identical models in a reduced theory can be used to represent physical situations for which we have a prior commitment to them being distinct in a way that the corresponding reduced models cannot.

Varieties of this point have been made previously. For one, there is a strand of literature on the empirical significance of symmetries that aligns closely with the idea that for representing *subsystems*, symmetry-related models can characterize empirically distinct situations.[9] Meanwhile, Belot (2018) and Fletcher (2020) argue that there is a sense in which isomorphic models can be said to generate distinct possibilities through the maps that relate them. Finally, the partial observables approach to gauge variables pioneered by Rovelli (2004) presents a picture under which symmetry-variant features of a theory can be representa-

[8]Dewar (2019, fn. 27) notes that his point is only that reduction has *some* explanatory deficiency, but the question of whether there is a stronger way to argue for sophistication still stands. See also Martens and Read (2021).

[9]See in particular Wallace (2022).

tionally useful. However, I take the novelty of my approach to be the focus on solving the following puzzle:

> **Puzzle**: How can a sophisticated theory have representational advantages compared to a reduced theory if they are theoretically equivalent, and so have the same content?

My resolution to this puzzle draws on a distinction between two kinds of structure that one can define in the context of a theory, which I call 'theoretical structure' and 'auxiliary structure' respectively.

Theoretical Structure is the structure that a theory attributes to the world in virtue of its "invariant" content; the content that is invariant under isomorphisms of the models of the theory. In other words, when models are isomorphic, they are equivalent in terms of theoretical structure. Moreover, one can exemplify the theoretical structure through the equivalence classes of the theory. This is indeed the standard way of explicating mathematical structure, and it is the structure that notions such as 'categorical equivalence' aim to capture.

Auxiliary Structure is the structure that one can define in the models of a theory in virtue of the way that one characterizes the theoretical structure. The auxiliary structure goes beyond the theoretical structure in that the mathematical tools one uses to characterize the theoretical structure may have further resources, such that one can differentiate more structure than simply the theoretical structure. In particular, it is only through auxiliary structure that one can talk about differences between isomorphic models. Of interest here is precisely the auxiliary structure that goes past the theoretical structure in this way.

To see this distinction in the handedness theory example, consider a model M of T_H consisting of two objects labelled a and b where $R^M = a$ and $L^M = b$. Now consider a second model N of the theory that consists of permuting the domain of the first model and pushing forward the properties such that $R^N = b$ and $L^N = a$. These two models agree on *theoretical structure* since they agree on how many objects have each handedness property. However, there is a difference between the models that one can describe using *auxiliary structure*: in model M object "a" is right-handed, while in model N, "a" is left-handed (and vice versa for object "b").

While reference to the domain of the models is one place where auxiliary and theoretical structure can diverge, the importance of these two kinds of structure in the handedness theory example arises when thinking about how they differ between the original, reduced, and sophisticated theory. In T_H, the theoretical structure and the auxiliary structure both include handedness structure i.e. structure through which one can define 'left' and 'right' hands as distinct. In T_R, neither the theoretical structure nor auxiliary structure include handedness structure; while one can say whether hands are congruent or not, one cannot say which are 'left' or 'right'.[10] In T_S, the theoretical structure and auxiliary structure diverge in a special way: the theoretical structure does not include handedness structure but the auxiliary structure does, since unlike in the reduced theory, one has access to the properties L and R in describing the models of the theory.

Usually, theoretical structure is understood to give the content of a theory and auxiliary structure is regarded as mere descriptional redundancy; it has to do with one's pragmatic choice of representation. On this understanding, isomorphic models have equivalent content and any differences between them described by auxiliary structure do not have any bearing on the representational capabilities of these models. However, I will argue against this view: auxiliary structure is not just descriptional redundancy but can have a representational role, and this is true even if one understands the theoretical structure to give the content of the theory. This will provide support for the claim that the sophisticated version of a theory

[10] In addition, one cannot reintroduce L and R in the reduced theory by fixing an element and saying any other element is L if it is congruent with the fixed element and R otherwise, since the theory doesn't allow one to identify a specific element.

can have representational benefits over the reduced version. I argue for this position through three claims:

> **Claim 1**: Isomorphic models can equally well represent a single physical situation in virtue of being equivalent in terms of theoretical structure.

This claim captures one made elsewhere that isomorphic models have the same "representational capacities" (Weatherall (2018); Fletcher (2020)). However, the emphasis here is on the role of theoretical structure – it is the theoretical structure that determines the extent to which models are able to represent some physical situation. The reason is that it is the theoretical structure that captures the physical content of the models of the theory, in the sense of being the structure attributed by the theory to the world.

The second claim is that there is an importantly different sense in which isomorphic models in a sophisticated theory can be physically distinguished:

> **Claim 2**: Auxiliary structure can be used to provide physically relevant distinctions between isomorphic models in a sophisticated theory.

In order to unpack this claim, let us return to the models of the sophisticated theory that consist of only 'left-handed' objects and only 'right-handed' objects respectively. One might argue that there is a natural sense in which they can represent distinct situations: they can represent two physical situations where the objects are in *different* congruence classes. But how is this compatible with the fact that these models have the same invariant structure that captures the physical content of the theory?

One response is to say that one could simply stipulate the interpretation of the models to be different, or, following Wallace (2022):

> "symmetry-related configurations can be understood as representing different possible configurations if we hold fixed the choice of representational convention." (p.337)

So in the above example, one might say that one could stipulate a standard of "left" across the models such that they represent different physical situations relative to this standard. But where does this representational convention come from, physically? In particular, since the theoretical structure is not sensitive to such choice, in what sense can one impose it on the models in order to distinguish the interpretation of these models?

Here is where auxiliary structure comes into play: the sense in which one can impose a physical representational convention across isomorphic models of a sophisticated theory, I argue, is that one can use the auxiliary structure of the theory, through which one can distinguish these models descriptionally, to represent further details about the physical situations represented by these models. When one can use auxiliary structure to represent an additional system that acts as a reference frame, one can distinguish the situations represented by isomorphic models that differ relative to the fixed auxiliary structure. For example, in the sophisticated models discussed above, one can use the auxiliary structure corresponding to the ability to define left and right handed objects to define a new object that is stipulated to be 'left-handed' in both models, such that relative to this hand, the models are distinct; one corresponds to the world where everything is congruent to this new hand, and the other corresponds to the world where everything is not congruent to the new hand. This use of auxiliary structure as defining a reference frame allows one to impose a representational convention that has a physical interpretation.[11]

[11] This argument bears several similarities with Fletcher (2020), who argues that we can understand the differences between isomorphic models in terms of the non-trivial maps that relate them. I take the view presented here to be to complementary to Fletcher's in that the reason auxiliary structure can be used to distinguish isomorphic models is related to the fact that the isomorphisms act non-trivially on this structure.

However, this would be of no interest if auxiliary structure could play the same role in the reduced theory. This leads to the final claim:

> **Claim 3**: Compared to the sophisticated theory, auxiliary structure in a reduced theory is unable to provide the same physically relevant distinctions between isomorphic models.

The reasoning is as follows. The reduction process equivocates between isomorphic models in the sophisticated theory that are non-isomorphic in the original theory; they correspond to identical models in the reduced theory. Therefore, while one could say that certain isomorphic models in the sophisticated theory correspond to distinct physical situations by using auxiliary structure to define a reference system, one cannot say the same of the corresponding identical models in the reduced theory.

To see this clearly, consider the same models as before and consider adding a new object that is handed in some way. We do not have the resources to stipulate that it represents a "left" or "right" hand in the reduced theory; we can only stipulate that it represents, for example, some hand congruent to all the hands in a model. But now, inasmuch as the models are identified in the reduced theory, the additional structure will not be able to distinguish these models by fixing its interpretation across them; if it is congruent to the hands in the first model, it is congruent in an identical model. Therefore, while physically relevant distinctions could be drawn between isomorphic models in the sophisticated theory using auxiliary structure, one cannot draw such distinctions in the corresponding reduced models.

To say that the additional hand is congruent to the objects in one model but not in an identical model, one would need to stipulate some additional fact about the difference between the two models, namely that all the hands in one model are in a different equivalence class to those of the other. But this fact can only be specified when talking about the models as subsets of a 'larger' model that includes the objects of both models, where the difference between the congruence of the objects can be specified in terms of *theoretical* structure. But this move is something that can also be made in the sophisticated theory; what is lacking in the reduced theory is the ability to distinguish these models without stipulating additional inter-model relations that rely on theoretical structure.

The combination of these three claims highlights that isomorphic models in a sophisticated theory play a multi-faceted role: they can be used to represent a single physical situation (Claim 1), and they can be used to represent distinct situations through a physical interpretation of the auxiliary structure as, for example, a fiducial system (Claim 2).[12] Inasmuch as this multi-faceted role is beneficial, Claim 3 demonstrates that the sophisticated version of a theory can have a representational advantage over the reduced version. Moreover, these three claims are compatible with the sophisticated and reduced theories being theoretically equivalent in the sense of attributing the same theoretical structure. We therefore have solved the puzzle stated earlier: there is no tension between a sophisticated theory having representational benefits while also being theoretically equivalent to a reduced theory.

This representational argument for sophistication is arguably more robust that the explanatory argument given by Dewar (2019), for the following reasons. First, one can point to precisely why a theory is superior if its models have the resources to represent a greater number of physical distinctions than the models of another theory: inasmuch as these physical distinctions are ones that one thinks a theory ought to capture, a theory whose tools prevent one from representing them is lacking. Second, unlike the fact that one can point to explanatory benefits of the reduced theory, there is not a corresponding representational benefit that can attributed to the reduced theory, inasmuch as reduction always constrains auxiliary structure. Therefore, the two worries raised previously for the explanatory account are bypassed.

[12]The fact that a physical interpretation is given to auxiliary structure does not imply that it should be promoted to theoretical structure. If one did so, one would no longer be able to use the models to represent a single situation, and one would return to the issue of a theory with excess structure.

However, the representational argument is aligned with the explanatory argument in the sense that part of the explanatory benefit of sophistication might be seen to come from its representational benefits. For example, consider the fact that two subsystems that each consist only of congruent hands can be such that their combined system does not consist only of congruent hands. This fact can be explained in the sophisticated theory through the difference in auxiliary structure between the representation of the subsystems: using auxiliary structure as a physical representational convention distinguishes the subsystems even though the models are equivalent in terms of theoretical structure. In the reduced theory, there are no relevant differences in auxiliary structure between the models representing these subsystems, and so the difference can only be given by asserting further statements about the relation between the subsystems. Therefore, one might argue, this fact falls out naturally from the sophisticated theory but must be stipulated the reduced theory. And so, the representational power of sophisticated theories lends itself to explanatory benefits.

3.4 Conclusion

We began with discussing the question of how to get rid of excess structure. The idea was that we should care only about what is invariant under certain symmetry transformations, and so we should remove any features of the theory that are not invariant. But in fact, we have seen that what varies under a symmetry transformation can be representationally useful, when it is part of the *auxiliary structure* of a theory. And so, while it might be true that what we care about when talking about *theoretical structure* are the features invariant under symmetry transformations, we have shown that what might be regarded as 'surplus' from this perspective need not be surplus from a wider perspective that includes the representational role of auxiliary structure.

In discussions on theory building and excess structure, it is often assumed that the only consideration is that of determining theoretical structure, where the way that one characterizes this structure is something that has only pragmatic value. This obscures the role of auxiliary structure and leads to confusion regarding cases where some structure appears 'surplus' and yet also seems to play an essential role. If we consider theory building to not just be about correctly characterizing theoretical structure, but also auxiliary structure, then this suggests that closer attention ought to be paid to the choices of auxiliary structure and the physical implications these choices have.

In light of this, arguments about excess structure *do* require consideration of whether one takes a theory to be a full description of all of the structure in the world: if one wants to allow for a theory to represent incomplete physical situations such as subsystems or situations relative to some reference frame, then certain kinds of auxiliary structure may be indispensable in a way that they are not if the theory is only used to represent the whole universe. We have discussed one example in this chapter; however, further work is necessary to characterize precisely the choices of auxiliary structure for some theory and the way that one associates representational aims with these choices.

Chapter 4

The Physical Significance of Partial Observables: Connecting Gauge and Surplus Structure

4.1 Introduction

In Chapter 3, I presented a division between two kinds of structure that I called *theoretical structure* and *auxiliary structure*, and I demonstrated that auxiliary structure can be used to play a representational role in the sophisticated version of a theory in a way that is absent in the reduced version of a theory. In this Chapter, I consider how this argument plays out in the context of the constrained Hamiltonian formalism.

In the constrained Hamiltonian formalism, the gauge variables are often regarded as the source of excess structure, since they are variables whose value is underdetermined by the dynamics. On the other hand, the gauge-invariant features, or the 'observables', are seen as the physically meaningful quantities. In light of this distinction, reduction is often argued to be the right approach to removing excess structure, since it redefines the theory in terms of the gauge-invariant quantities.

However, it is also well-known that reduction in the constrained Hamiltonian formalism can have puzzling conclusions. Most notably, for theories that are time-reparameterization invariant, reduction leads to a strong version of the ‘Problem of Time’: there is no meaningful notion of evolution on the reduced space since the points of the reduced space are entire solutions. This suggests that closer consideration ought to be paid to the kind of redundancy that is being characterized by the notion of a gauge variable.

One influential approach that argues that the gauge variables should not be removed from a theory is called the ‘partial observables’ program, pioneered by Carlo Rovelli. In this Chapter, I connect Rovelli’s approach to the arguments I gave in Chapter 3 by showing that one can reconstruct the partial observables approach as presenting a view according to which sophistication in the context of the constrained Hamiltonian formalism comes with representational benefits.

In more detail, I will argue that the literature regarding the interpretation of gauge transformations conflates the distinction between theoretical structure and auxiliary structure, and that this leads to confusion in the literature regarding the role of gauge variables. I will then argue that one can understand the partial observables program as providing an account of the representational role that a particular kind of auxiliary structure plays, and that this motivates a sophisticated version of the constrained Hamiltonian formalism.

4.2 Argument for Reduction

In Dirac’s presentation of the constrained Hamiltonian formalism, there is a clear sense in which there is arbitrariness when there are first-class constraints present: the dynamics are defined in terms of the Total (or Extended) Hamiltonian, which includes arbitrary functions in their definition.[1] This means that given some initial state (q_0, p_0), there can be multiple possible evolutions and multiple possible values for a dynamical variable at later times. As Dirac (1964) presents it, this is a kind of *indeterminism* in the mathematical formalism that doesn't correspond to any physical indeterminism: the multiple evolutions from some fixed initial state should be regarded as physically equivalent because they arise from arbitrariness in the Hamiltonian.

An alternative way to present the arbitrariness in the constrained Hamiltonian formalism is in terms of *underdetermination*, which is seen most naturally in the geometric way of presenting the constrained Hamiltonian formalism on the constraint surface: there is an underdetermination of states and of solutions arising from the fact that the induced two-form has null vector fields. In particular, states along the gauge orbits cannot be distinguished by the induced two-form, and solutions are only defined up to vector fields associated with the first-class constraints.[2]

In light of such arbitrariness in the constrained Hamiltonian formalism, it is tempting to make the following interpretive move: one should interpret gauge-related states/solutions, and in turn, variables that vary between gauge-related states/solutions, as being equivalent.[3] Therefore, only the quantities that are gauge-invariant (the "observables") should be interpreted as physically meaningful, while the gauge-dependent quantities (the "gauge variables") should be interpreted as not having physically meaningful values, since their value varies between states/solutions that are regarded as equivalent.

[1] cf. Section 1.2.

[2] cf. Section 1.6.

[3] One can distinguish the view that gauge-related *states* should be identified and the view that gauge-related *solutions* should be identified. Indeed, these views might lead to importantly distinct options both mathematically and conceptually. However, for the purposes here, these two views will be considered together. The reason is that in the geometric way of presenting the constrained Hamiltonian formalism, the same transformations generate state gauge-transformations and solution gauge-transformations, and so at least mathematically the same identification takes place. For further discussion of the conceptual differences, see, for example, Wallace (2002) and Gryb and Thébault (2023).

However, there are arguments in the literature suggesting that merely making this interpretational move is not sufficient to remove the arbitrariness present in the constrained Hamiltonian formalism; rather, one should change the mathematical formalism itself such that gauge-related states/solutions are identified. Recall from Chapter 2 that we can reformulate a constrained Hamiltonian theory as a reduced theory: a theory whose state space is given by the reduced version of the constraint surface and whose dynamics are given in terms of a Hamiltonian function and a symplectic two-form defined on the reduced space. Several authors have argued that only the move to the reduced phase space is sufficient to remove the arbitrariness present in the constrained Hamiltonian formalism. For example, Thébault (2012) says that one could interpret a theory in the constrained Hamiltonian formalism by "instituting a many-to-one relationship between gauge related sequences of points on the constraint surface and the unique sequences of instantaneous states they represent" but that this "does nothing about removing what would seem like superfluous mathematical structure – to dispense with this surplus structure we need to move to the reduced phase space." Similarly, Belot (2003) says that "because the points of such orbits are dynamically indifferent, the x_i [the gauge variables] are dynamically irrelevant – any way of setting their value leads to the "same" evolution. This suggests in turn that it may be possible to drop the x_i from our theory altogether".[4] Both authors agree that it is not enough to interpret the gauge-dependent variables in such a way that trajectories on which they differ are regarded as equivalent – one must remove them from the theory in order to remove the redundancy present. The reduced theory provides such an option since the points of the reduced space are just the equivalence class of points along the gauge orbits.

Let us call the following the **Reduction Argument**:

> (1) States and solutions related by gauge transformations are mathematically inequivalent in the constrained Hamiltonian formalism.

[4]For another expression of this view, see Belot and Earman (2001).

(2) States and solutions related by gauge transformations are physically equivalent, and so the gauge variables should not be regarded as physically significant.

(3) In an ideal theory, mathematical equivalence should align with physical equivalence and there should be no definable quantities that lack physical significance.[5]

(Conclusion) Therefore, we should remove the gauge variables and the differences between gauge-related states/solutions by reformulating the theory on the reduced space.

In the rest of this Chapter, we will consider whether this argument holds up. In particular, there are two parts of the argument that need further clarification. First, to say whether certain states/solutions are mathematically inequivalent, we need an account of mathematical equivalence in the context of the constrained Hamiltonian formalism. There is a naive sense in which gauge-related states/solutions are mathematically inequivalent: they correspond to/are composed of different points of the state space. However, mathematical equivalence is usually cashed out in terms of *isomorphism*. Whether a non-trivial transformation on state space gives rise to mathematically inequivalent states depends upon whether it is an isomorphism or not, which in turn depends on the mathematical structure that we take to represent the constrained Hamiltonian formalism.

Second, an important step in this argument is the claim that the gauge variables are not *physically significant* on the basis that their value differs between gauge-related states/solutions. In order for this claim to have the consequence that one ought to remove the gauge variables from the theory by moving to the reduced phase space, it must be that a gauge variable not having physical significance means that it plays no role in a theory. However, merely differing in value between points/curves that one takes to represent physically equivalent

[5]These two things – alignment between mathematical and physical equivalence, and no quantities without physical significance – are interrelated. However, we will see that they can come apart depending on what one means by *physical significance*.

states/evolutions does not necessarily imply that it cannot be used to represent a physical quantity. Therefore, there is room to argue that although the value of a gauge variable is underdetermined by the theory, one should not remove gauge variables from the theory altogether. The most influential response of this kind is the partial observables program, to which we now turn.

4.3 Partial Observables Approach to Gauge

The partial observables approach was pioneered by Carlo Rovelli as a way to resolve the Problem of Time, as well as explain the ineliminable role that gauge variables seem to play more generally.[6] The basic idea of the partial observables approach is the following: gauge-dependent quantities, despite not being *predictable* within a theory (since their value is underdetermined), are still important because it is the couplings between certain gauge-dependent quantities (the 'partial observables') that give rise to the gauge-independent quantities (the 'complete observables'). Therefore, partial observables play a role in providing the physical content of a theory through their role in making up the gauge-invariant quantities. In addition, one can measure the partial observables when they couple by using the invariant relationship to determine the value of one partial observable when another takes a fixed value. In this way, they can be "measured but not predicted" (Rovelli (2014)). Consequently, partial observables are not redundant to a theory.

To motivate this view, let us consider an example adapted from Rovelli (2014). Take a system of two spaceships in Euclidean space represented by the Lagrangian:

[6] See, for example, Rovelli (2002, 2014, 2004). Extensions to Rovelli's approach have been given, for example, by Dittrich (2006, 2007); Thiemann (2008).

$$L_x = \frac{1}{2}(\dot{x}_2 - \dot{x}_1)^2 \tag{4.1}$$

Moving to the Hamiltonian framework, we have that $p_1 = \dot{x}_1 - \dot{x}_2$ and $p_2 = \dot{x}_2 - \dot{x}_1$. This gives rise to a single primary constraint $\phi_1 \approx 0$ where $\phi_1 = p_1 + p_2$. The Total Hamiltonian is given by:

$$H_T = \frac{1}{2}(p_2)^2 + \mu(p_1 + p_2) \tag{4.2}$$

where μ is arbitrary. There are no secondary constraints since $\{\phi_1, H_T\} = 0$. This implies that ϕ_1 is first-class and generates a gauge transformation, meaning that the evolution of the system is underdetermined. In particular, we find that x_1 and x_2 are not observables, while $x_2 - x_1$ is an observable. This suggests if we follow the Reduction Argument that one should remove the quantities x_1 and x_2 from the theory and rewrite the theory in terms of the quantity $a = x_2 - x_1$ via $H_R = \frac{1}{2}p_a^2$, which has no gauge freedom.

Consider also a second system composed of two spaceships, represented by the Lagrangian:

$$L_y = \frac{1}{2}(\dot{y}_2 - \dot{y}_1)^2 \tag{4.3}$$

The Hamiltonian analysis of this system is the same; we can rewrite the theory in terms of the quantity $b = y_2 - y_1$ via $H_R = \frac{1}{2}p_b^2$ in order to remove the gauge freedom.

But now consider bringing the two systems together so that the combined system is represented by the Lagrangian:

$$L_c = \frac{1}{2}(\dot{x}_2 - \dot{x}_1)^2 + \frac{1}{2}(\dot{y}_2 - \dot{y}_1)^2 + \frac{1}{2}(\dot{y}_1 - \dot{x}_2)^2 \tag{4.4}$$

where we have an interaction term that goes past simply the sum of the two individual systems' Lagrangian. In the Hamiltonian analysis, we find that the four canonical momenta give rise to a single primary constraint $\phi_c \approx 0$ where $\phi_c = p_{x_1} + p_{x_2} + p_{y_1} + p_{y_2}$. The Total Hamiltonian is given by:

$$H_c = \frac{1}{2}p_{x_1}^2 + \frac{1}{2}p_{y_2}^2 + \frac{1}{2}(p_{y_1} + p_{y_2})^2 + \mu(p_{x_1} + p_{x_2} + p_{y_1} + p_{y_2}) \tag{4.5}$$

Again, none of the individual variables x_1, x_2, y_1, y_2 are observables, while $x_2 - x_1$ and $y_2 - y_1$ are. However, we also have that $y_1 - x_2$ and all other such combinations of x and y variables are also observables. This joint system therefore has more gauge-invariant quantities than the sum of the two individual systems.

The fact that one has more gauge-invariant quantities than the sum of the two individual systems leads Rovelli (2014) to conclude that gauge variables are crucial when systems interact, since one could not form this combined system using just the gauge-invariant quantities of the individual systems. In other words, if one 'got rid' of the gauge variables when modelling the individual systems by defining these systems only in terms of the invariant quantities a and b, one wouldn't be able to describe the coupled system without introducing a new variable that is defined in terms of the relation between the gauge variables of the individual systems, since one cannot define the quantity $\dot{y}_1 - \dot{x}_2$ (or similarly $p_{y_1} + p_{y_2}$) merely in terms of a and b.

Rovelli (2002) argues that something similar can be done to avoid the Problem of Time. One way of stating the problem is that when the Hamiltonian is itself a first-class constraint, there is no change over time for any physical variable since the observables are invariant under time evolution. Rovelli argues that this rests on confusion regarding the notion of evolution: there is evolution in coordinate time, and there is physical evolution. Physical evolution can be maintained if one drops the requirement that evolution must be in coordinate time, and rather takes evolution to be a relative notion between partial observables. For example, in General Relativity, proper times depend on the coordinate time and therefore are gauge variables; they are not invariant under time evolution. However, the relative evolution between two different proper times representing two different observers is a complete observable, since it is independent of coordinate time and so is invariant under evolution in coordinate time. Therefore, we can regard proper times as being partial observables in the same way that the positions of the spaceships were partial observables: there is a complete observable that can be understood as the relation between these partial observables. If one considers this observable to define physical evolution, then physical change does not disappear, even though one does not have an independent time coordinate.

These examples, despite giving insight into Rovelli's program, still leave open some philosophical questions. The crucial question of importance for the purposes here is what the physical significance of the partial observables is meant to be, in contrast to the complete observables. Rovelli (2002) says that "in a sense, they are the quantities with the most direct physical interpretation in the theory". However, their physical interpretation needs to be squared with the fact that the partial observables are not *predictable*, in the sense that their value is not fixed by the theory. Rovelli suggests that this tension is resolved through the fact that partial observables are still "measurable". However, one might respond that the value of a partial observable is only determined through a measurement of a complete observable that results from the coupling of partial observables, and thus it is ultimately only the complete observables that are measurable.[7] This suggests that one cannot rely on the measurability of partial observables to explain their physical significance. Moreover, the fact that partial observables are not predictable means that one cannot think of a theory as providing a full account of them since the theory does not specify their dynamics uniquely. So we are left with the question: how can a quantity that is not fully specified by a theory have a direct physical interpretation in that theory?

A different way of phrasing this question is the following: how does the partial observables program resolve the fact that the constrained Hamiltonian formalism appears to have *excess structure* in the sense given by the Reduction Argument, namely, that gauge transformations relate mathematically inequivalent but seemingly physically equivalent situations? Although the partial observables program provides an account under which gauge variables can be physically meaningful, it doesn't say that the situations related by a gauge transformation are physically distinct: such a transformation changes the value of the gauge variables, but not their relative value, and so inasmuch as one measures the complete observables a gauge transformation doesn't give rise to a physically distinct situation. Therefore, the partial observables program seems to have a tension between the physical role attributed to partial observables and the apparent excess structure in the theory.

4.4 The Physical Significance of Partial Observables

In this section, I will argue that the tension between the physical interpretation of partial observables and their lack of full specification can be resolved if we understand the partial observables program in terms of the distinction between *theoretical structure* on the one

[7]This is the view taken by Thiemann (2008).

hand, and *auxiliary structure* on the other hand, and the kind of physical significance that can be attributed to both. In doing so, I will provide my own account of why the Reduction Argument fails.

Recall from Chapter 3 that the distinction is as follows:

Theoretical Structure is the structure that a theory attributes to the world in virtue of its "invariant" content; the content that is invariant under isomorphisms of the models of the theory. In other words, when models are isomorphic, they are equivalent in terms of theoretical structure. Moreover, one can exemplify the theoretical structure through the equivalence classes of the theory. This is indeed the standard way of explicating mathematical structure, and it is the structure that notions such as 'categorical equivalence' aim to capture.

Auxiliary Structure is the structure that one can define in the models of a theory in virtue of the way that one characterizes the theoretical structure. The auxiliary structure goes past the theoretical structure in that the mathematical tools one uses to characterize the theoretical structure may have further resources, such that one can differentiate more structure than simply the theoretical structure. In particular, it is only through auxiliary structure that one can talk about the differences between isomorphic models. Of interest here is precisely the auxiliary structure that goes past the theoretical structure in this way.

In Chapter 3, I argued that auxiliary structure can play a representational role, and that this provides an argument in favor of *sophistication* over *reduction* since auxiliary structure in the sophisticated theory can be lost when moving to the reduced theory.

In order to use the distinction between theoretical structure and auxiliary structure in the context of the constrained Hamiltonian formalism, we need to spell out the models and symmetries of the theory.[8] More precisely, we are going to characterize different formulations

[8]Here we follow the category-theoretic perspective of characterizing theories, where theories are not just given as a collection of models or a collection of sentences; rather, they come *structured*, which means that

of the theory as *categories of models*, where we specify not only the objects of the category as models of the theory, but also specify the isomorphisms of the theory as (invertible) arrows of the category. There are many ways to do this.

First, we could define the objects of the theory as models (T^*Q, H, φ_i), a cotangent space equipped with a Hamiltonian function and a collection of constraints. The equations of motion would then be Hamilton's equations $\omega(X_H, \cdot) = dH$ along with the constraints $\varphi_i = 0$, which tell one the collection of dynamically-allowed points of cotangent space. The arrows of such a theory between models (T^*Q, H, φ_i), (T^*Q, H', φ_i') are naturally given by point*-transformations that preserve H and φ_i. Recall from Chapter 2 that the point*-transformations T^*f are such that given by diffeomorphism $f : T^*Q \to T^*Q$, $T^*f : (q, p) \to (f^{-1}(q), f^*(p))$. So preserving H and φ_i just means that $H' \circ T^*f = H$ and $\varphi_i' \circ T^*f = \varphi_i$.

Alternatively, we might want to take the state space to be given more generally by a symplectic manifold, such that the models consist of $(T^*Q, \omega, H, \varphi_i)$. Then, the arrows between models $(T^*Q, \omega, H, \varphi_i)$, $(T^*Q, \omega', H', \varphi_i')$ are naturally taken to be symplectomorphisms that also preserve H and φ_i. That is, f is a symmetry between models if $f : T^*Q \to T^*Q$ is a diffeomorphism such that $f^*(\omega') = \omega$, $f^*(H') = H$ and $f^*(\varphi_i') = \varphi$.

The difference between these two options – whether we take the state space to be given by the structure of cotangent space or a symplectic manifold – is not crucial here. But since all point*-transformations are symplectomorphisms, we will focus on the second option.[9] Let us call this theory **ConHam1**. Importantly, arbitrary gauge transformations (transformations generated by arbitrary combinations of the first-class constraints) do not in general correspond to symmetries of the models of **ConHam1**.

they also come along with a standard of equivalence. For more discussion, see in particular Halvorson (2012, 2016); Weatherall (2016a, 2019a,b).

[9] See Barrett (2015b) for proof that this only goes in one direction; not all symplectomorphisms are point*-transformations.

Proposition 4.1: Gauge transformations are not, in general, arrows in **ConHam1**.[10]

.

This shows that if we take the structure of constrained Hamiltonian mechanics to be given by **ConHam1**, then there is a sense in which gauge variables are part of the theoretical structure: models that differ regarding their value are distinct according to the theory.

However, given that gauge variables appear to be 'excess structure' for the reasons given earlier, it seems natural that we might want to formulate a theory such that gauge variables are not part of the theoretical structure. In particular, we find that on the above characterization, the dynamics are such that the Hamiltonian function H and the function $H + v^a\varphi_a$ are equivalent: they correspond to equivalent trajectories, once we take into account the constraints. Therefore, we want a way of writing down the theory that renders these Hamiltonians equivalent.

We might interpret this as saying that we want models $(T^*Q, \omega, H, \varphi_i)$ and $(T^*Q, \omega, H + v^a\varphi_a, \varphi_i)$ to be symmetry-related. But this runs into difficulties. For one, it is not clear how to define this symmetry as an action on phase space, inasmuch as gauge transformations are not symplectomorphisms. Second, it isn't clear that this really captures what we are after, since although it categorizes models with different evolutions as equivalent, we might want to characterize the claim that particular *states* are equivalent.

Alternatively, we might say that models $(T^*Q, \omega, H, \varphi_i)$ and $(T^*Q, \omega', H', \varphi_i)$ are symmetry-related when a diffeomorphism $f : T^*Q \to T^*Q$ is such that $f^*\omega'_c = \omega_c$ and $f^*H'_c = H_c$ where ω_c and H_c represent the two-form and Hamiltonian restricted to the constraint surface. That is, f is a symmetry when it preserves the structures on the constraint submanifold. The reason for this definition is that the constraint submanifold is taken to represent the dynamically

[10] See C.1 for proof.

allowed states, and so if the structures on the constraint submanifold are preserved, then this would naturally give rise to the same physical situation. But while such an option improves on the former in that we can both understand the symmetry as an action on phase space and it provides a sense in which certain states are equivalent, there is something still unnatural about it. In particular, inasmuch as what one cares about is the preservation of structure on the constraint surface, it is not clear what role the points outside of the constraint surface are playing.

Instead, it seems that the natural way to define the theory would be to define the models as $(\Sigma_f, \tilde{\omega}_f, H)$ where Σ_f is the final constraint surface, $\tilde{\omega}_f$ is the presymplectic two-form on Σ_f, and H is the Hamiltonian restricted to the points of Σ_f, with arrows as symplectomorphisms that preserve H.[11] Call this **ConHam2**.

Proposition 4.2: Arbitrary gauge transformations are arrows in **ConHam2**.[12]

The upshot is that in **ConHam2**, models that differ regarding the value of the gauge variables are equivalent according to the theory. That is, if we consider models to have additional structure corresponding to a collection of gauge variables A_i, then models $(\Sigma_f, \tilde{\omega}_f, H, A_i)$, $(\Sigma_f, \tilde{\omega}_f, H, A_i')$ where A_i' are the gauge transformed version A_i are equivalent according to **ConHam2**, even though A_i' and A_i differ in value at the same point on Σ_f. Indeed, I think that it is natural to regard **ConHam2** as capturing the *(internally) sophisticated* theory: it is a theory where the isomorphisms correspond precisely to the transformations that we regard as symmetries of the theory, which includes the gauge transformations.[13]

[11] Recall that this is how we defined the Extended Hamiltonian theory in Section 1.8.1.

[12] See C.2 for proof.

[13] There is a disanalogy with sophistication as presented in Dewar (2019) since the move from **ConHam1** to **ConHam2** not only alters the isomorphisms of the theory, but it also alters the state space by removing the points off the constraint surface. I comment on this further in the Conclusion.

Another way to remove the 'excess structure' coming from the gauge variables is to define the models on the reduced space. This would be to take the models to consist of $(\bar{\Sigma}, \bar{\omega}, \bar{H})$ with symmetries being symplectomorphisms that preserve $\bar{H}$. Call this **ConHamRed**

> **Proposition 4.3**: Arbitrary gauge transformations are arrows in **ConHamRed**.[14]

We therefore have that different ways of characterizing the models and symmetries of a constrained Hamiltonian formalism lead to different conclusions regarding whether some quantity is part of the theoretical structure or not i.e. whether it is agreed upon under the isomorphisms of the theory or not. If one takes **ConHam1** to represent constrained Hamiltonian theories, then gauge variables are part of the theoretical structure, since there are models where their value differs that are distinct according to the theory. However, if one takes **ConHam2** or **ConHamRed** to represent constrained Hamiltonian theories, then the gauge variables are not part of the theoretical structure, since models related by a gauge transformation are *isomorphic*, and recall that the theoretical structure is given by structure preserved under isomorphism.

Turning now to auxiliary structure, it is clear that in **ConHam1**, gauge variables count as auxiliary structure, inasmuch as auxiliary structure always includes the theoretical structure. More interesting is that gauge variables are auxiliary structure in **ConHam2**, but not in **ConHamRed**. The reason that gauge variables are auxiliary structure in **ConHam2** is just that they are well-defined functions on Σ_f: they are just the functions whose value varies along the gauge orbits. The reason that gauge variables are not auxiliary structure in **ConHamRed** is that one cannot define the gauge variables as functions on $\bar{\Sigma}$, since by definition $\bar{\Sigma}$ equivocates between points where the value of gauge functions differ.

[14] See C.3 for proof.

It is these distinctions between **ConHam2** and **ConHamRed** that I think have been conflated in the literature. Let us return to the Reduction Argument. Premise 1 of the argument was:

> (1) States and solutions related by gauge transformations are mathematically inequivalent in the constrained Hamiltonian formalism.

We can now see that there is an ambiguity in this premise since there are different ways to characterize the models and symmetries of the constrained Hamiltonian formalism that lead to different conclusions regarding whether gauge transformations are mathematically inequivalent. In particular, if we take the theory to be **ConHam2** i.e. the theory formulated on the final constraint surface, then gauge transformations are symmetries of the theory, and therefore states/solutions related by gauge transformations are naturally regarded as mathematically *equivalent* in this theory.

There are also ambiguities with the other premises of the argument:

> (2) States and solutions related by gauge transformations are physically equivalent, and so the gauge variables should not be regarded as physically significant.

If we take for granted the first clause of the premise, then one can infer that gauge variables are not physically significant in the sense of not being structure we want our theory to attribute to the world through the theoretical structure. However, one cannot infer that gauge variables play no physically significant role as *auxiliary structure* without some further claim that gauge variables do not have any representational role.[15] Therefore, this premise doesn't tell us on its own whether gauge variables should be part of the auxiliary structure of a theory or not.

[15]cf. Chapter 3

(3) In an ideal theory, mathematical equivalence should align with physical equivalence and there should be no definable quantities that lack physical significance.

In light of the distinction between theoretical and auxiliary structure, we can see that the two parts of this premise come apart: mathematical equivalence and physical equivalence can be aligned while it also being the case that there are definable quantities (from auxiliary structure) that lack physical significance, in the sense of not being structure we want our theory to attribute to the world through *theoretical structure*. In particular, we have argued that one may want definable quantities that lack physical significance in terms of theoretical structure but do have physical significance in terms of auxiliary structure.

(Conclusion) Therefore, we should remove the gauge variables and the differences between gauge-related states/solutions by reformulating the theory on the reduced space.

From the above, we can see that this conclusion only follows if we take gauge variables to lack physical significance as part of *auxiliary structure* as well as as part of theoretical structure, since the reduced theory changes the auxiliary structure of the theory by removing the ability to define the gauge variables. If we take gauge variables to play a role as part of auxiliary structure, then the argument only shows that we ought to formulate a theory such that gauge variables are not part of the theoretical structure. But we have a way to do this that doesn't involve moving to the reduced theory: the sophisticated option given by **ConHam2**. This formulation allows one to remove gauge variables from the theoretical structure without removing them from the auxiliary structure.

To summarize, filling out the ambiguities of the Reduction Argument gives us the following:

(1) States and solutions related by gauge transformations are mathematically inequivalent in the theory characterized by **ConHam1**.

(2) States and solutions related by gauge transformations are physically equivalent, and so the gauge variables should not be part of the theoretical structure of the theory.

(3) In an ideal theory, mathematical equivalence should align with physical equivalence and there should be no definable quantities that play no representational role (as part of theoretical or auxiliary structure).

(Conclusion) Therefore, we should remove the gauge variables and the differences between gauge-related states/solutions from **ConHam1** by reformulating the theory as **ConHamRed**.

On this framing, we can see that the argument is not valid without a further premise:

(4) Gauge variables play no representational role as part of auxiliary structure.

It is this premise that one can interpret the partial observables program as arguing is false, such that the Reduction Argument is not sound. In other words, one can reconstruct the partial observables program as showing that the Reduction Argument is misguided precisely because gauge variables can play a physical role as part of the auxiliary structure, and therefore they should not be removed by moving to the reduced theory. Their role as auxiliary structure is that they represent a system that can be measured through its coupling with other partial observables, and which combine with other partial observables to form the complete observables. This is not in conflict with the claim that partial observables should not be part of the theoretical structure; indeed, one might view this as being the crucial difference between the partial and complete observables.

In light of such a view, we can see that the conclusion that the Reduction Argument allows us to reach is only the following:

> (Conclusion, revised) We should remove the gauge variables and the differences between gauge-related states/solutions from **ConHam1** by reformulating the theory as **ConHam2**.

In other words, we can see the partial observables program as providing an argument in favor of *sophistication*, by the lights of the premises of the Reduction Argument: commitment to premises 1 through 3 and the negation of premise 4 leads to the revised conclusion, which is precisely to say that the right formulation of the theory is the sophisticated one, rather than the reduced one.

Although this is not the characterization that Rovelli gives of the partial observables program, I think it helps to make sense of the claim that partial variables are "measurable but not predictable". The sense in which they are not predictable is just that their value is underdetermined by the theory, and so we should formulate the theory such that models where the value of the partial observables differ ought to be isomorphic i.e. they should not be part of the theoretical structure. The sense in which they are measurable is that they can be used as auxiliary structure to represent quantities whose value can be determined when coupled with other quantities.

To further push the point, let us connect the argument made in Chapter 3 regarding the handedness theory example and the spaceship example that Rovelli gives. In the handedness theory, the role played by auxiliary structure was that it could act as a reference system relative to which one can distinguish the situations represented by isomorphic models. But there is another way that we can view the situation. Take the models related by the fact that the handedness of each object is 'swapped'. Instead of taking the reference to 'left' or 'right' to fix some auxiliary structure relative to which these models are distinguished, let us imagine that these models represent two subsystems of some larger system. Then, the relative handedness between objects across the models becomes important, inasmuch as it leads to different physical situations represented by the larger system. Moreover, one cannot represent the sense in which these models represent subsystems with different relative handedness in the reduced theory, since the reduced versions of these models are identical: there is no auxiliary structure that distinguishes the reduced models. To capture the differences in the reduced theory, one would have to stipulate additional information about whether objects across the models are congruent or not, which changes the description of the situation. But this is just what Rovelli argues is the case in the spaceship example: if we reduce the theories representing the individual systems, then we cannot talk about the coupling between them without introducing a new variable. On the other hand, in the unreduced theories, there is a way to describe the coupling in terms of auxiliary structure: the coupling is given by the relation between gauge variables.

This suggests that the spaceship example can be seen as picking out the same kind of representational significance as that highlighted in Chapter 3: auxiliary structure can be stipulated to represent some physical structure that can be used to distinguish physical situations modelled by the theory. The partial observables approach can be seen as providing a way to pick out the auxiliary structure that plays this role within the constrained Hamiltonian formalism: the partial observables are a collection of functions definable on phase space whose value is freely specifiable, but whose relationships with one another jointly suffice to characterize the complete observables.

There is a sense in which this view might be seen to deflate the role of partial observables: it takes partial observables to still be 'excess structure' in the theory given by **ConHam1**, since they distinguish non-isomorphic but symmetry-related models. Moreover, this view takes the physical role of partial observables to be dependent on the kinds of situations that a practitioner wants to represent. That is, partial observables are taken to only be physically meaningful inasmuch as they have a practical role in modelling subsystems and their relationships or additional physical systems that aren't contained in the theoretical structure; they are not taken to be structure that the theory itself attributes to the world. However, I take this flexibility to be a benefit of the account: it highlights that consideration needs to be paid to the use and interpretation of a theory in characterizing the theoretical and auxiliary structure of a theory.

4.5 Conclusion

In this Chapter, I have argued that we can reconstruct the partial observables program as providing an argument in favor of *sophistication* in the context of the constrained Hamiltonian formalism. This view reconciles the apparent tension that the partial observables program considers gauge variables to both not be predictable within a theory and to play an ineliminable role, since it highlights that gauge variables can be removed from the theoretical structure of the theory while still playing a role as part of the auxiliary structure. The notion of a partial observable provides a way to discriminate the gauge variables that play precisely this role.

While the focus in this chapter was on the way that one can motivate the partial observables program, exploring this issue highlighted several other interesting features of the constrained Hamiltonian formalism. Importantly, it highlighted that the theory formulated on the final constraint surface essentially comes "sophisticated": gauge transformations are isomorphisms of the mathematical structure of the final constraint surface. Therefore, on the understanding that excess structure comes from the presence of non-isomorphic but physically equivalent models, the theory formulated on the final constraint surface does not have any excess structure. However, the way we got to the sophisticated theory was not by starting with the unsophisticated theory formulated on cotangent space T^*Q (**ConHam1**) and using new mathematical/semantic tools such that the non-isomorphic but symmetry-related models are rendered isomorphic. Instead, we get to the sophisticated theory from **ConHam1** by simply removing certain points from T^*Q and thinking about the structures projected to this submanifold (with the same notion of isomorphism as **ConHam1**). This might in fact be seen as a form of *reduction* since it effectively equivocates between all models that vary only outside of the constraint surface.

One way to understand the difference between the constrained Hamiltonian formalism and the cases that Dewar (2019) considers is in terms of the role that constraints play in the kinematics vs. dynamics of the theory. In **ConHam1**, the constraints pick out the dynamically accessible points of the cotangent space, and therefore the dynamically possible models lie only on a submanifold of the state space of the kinematically possible models. On the other hand, in **ConHam2** the constraints are effectively built into the kinematically possible models, and so the dynamically possible models have the same state space as the kinematically possible models. Indeed, in **ConHam2**, the dynamically possible models arguably coincide with the kinematically possible models. Therefore, removing excess structure involves changing the relationship between the kinematical and dynamical state space. Indeed, one way of thinking about the 'excess structure' contained in **ConHam1** is just that it treats as kinematically important points of cotangent space that play no role in the physical content of the theory. On the other hand, in cases that Dewar (2019) considers, the change from the original theory to the sophisticated theory keeps the kinematics/dynamics division the same. This suggests that the way that reduction and sophistication are distinguished in Dewar (2019) in terms of how they are constructed from some unsophisticated/unreduced does not capture all cases that we might characterize as examples of 'sophistication' and 'reduction'.

A second difference with the cases that Dewar discusses arises when considering the question of whether the sophisticated and reduced versions of a theory are *equivalent*. Dewar (2019) argues via examples that the sophisticated and reduced versions of a theory will be *categorically* equivalent. In the case of **ConHam2** and **ConHamRed**, one might be inclined to think that these theories are equivalent, since one constructs **ConHamRed** by equivocating between mathematically equivalent state space points of objects in **ConHam2**. However, non-trivial gauge transformations correspond to non-trivial automorphisms of the models of **ConHam2**, and these transformations act trivially on the corresponding reduced model. This suggests that any suitable functor between the categories will not be injective on arrows. More precisely, take the functor F that takes the object $(\Sigma_f, \tilde{\omega}_f, H)$ to $(\bar{\Sigma}, \bar{\omega}, \bar{H})$ and that takes an arrow $g : \Sigma_f \to \Sigma_f$ to $\bar{g} : \bar{\Sigma} \to \bar{\Sigma}$ where $\bar{g}(\bar{x}) = g(\bar{x})$ i.e. $\bar{g}$ is the action of g on the gauge orbits, such that whenever g acts by moving points of Σ_f along the gauge orbits, $\bar{g}$ acts as the identity on $F(\Sigma_f)$. Then, the following is true:

> **Proposition 4.4:** F : **ConHam2** $\to$ **ConHamRed** is full and essentially surjective but not faithful i.e. F forgets *stuff*.[16]

This suggests that the difference between **ConHam2** and **ConHamRed** is not merely that they differ in auxiliary structure; they also differ in terms of how many ways a model is equivalent to itself. The significance of this property is discussed in Bradley and Weatherall (2020); we suggest that theories with less *stuff* are often associated with more (theoretical) *structure*. Whether this is the right interpretation of what is going on in the case of **ConHam2** and **ConHamRed** is something I hope to consider in future work.

There are two further sets of questions that this chapter raises. First, we saw in Chapter 2 that there is a many-to-one relationship between points of the Lagrangian final constraint surface and points of the Hamiltonian final constraint surface. This means that there are functions that one can define on the final Lagrangian constraint surface that are not projectable to the Hamiltonian final constraint surface, namely, those that vary between the points that are equivocated when moving to the Hamiltonian constraint surface. Are such functions plausible candidates for partial observables? And if so, does this provide a way

[16] See C.4 for proof.

to distinguish the interpretation of Lagrangian and Hamiltonian mechanics formulated on the final constraint surface, even if there is a sense in which they are structurally the same? Second, what does the sophisticated theory look like in the *quantized* version of a constrained Hamiltonian theory? Can we reconstruct the distinction between a reduced and sophisticated theory in a natural way, and how does it relate to the classical versions? These questions will also be left for future work.

Bibliography

Abraham, Ralph and Jerrold E. Marsden (1987). *Foundations of Mechanics, Second Edition*. Addison-Wesley Publishing Company, Inc.

Anderson, Edward (2012). Problem of time in quantum gravity. *Annalen der Physik*, 524(12): 757–786.

Baez, John, Toby Bartel, and James Dolan (2004). Property, structure, and stuff. Available at: `https://math.ucr.edu/home/baez/qg-spring2004/discussion.html`.

Baker, David John (2010). Symmetry and the metaphysics of physics. *Philosophy Compass*, 5(12): 1157–1166.

Barbour, Julian B. (1994). The timelessness of quantum gravity: I. the evidence from the classical theory. *Classical and quantum gravity*, 11(12): 2853.

Barbour, Julian B. and Brendan Z. Foster (2008). Constraints and gauge transformations: Dirac's theorem is not always valid. *arXiv preprint arXiv:0808.1223*.

Barrett, Thomas William (2015a). On the structure of classical mechanics. *The British Journal for the Philosophy of Science*, 66(4): 801–828.

Barrett, Thomas William (2015b). Spacetime structure. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, 51: 37–43.

Barrett, Thomas William (2019). Equivalent and inequivalent formulations of classical mechanics. *The British Journal for the Philosophy of Science*, 70(4): 1167–1199.

Batlle, C., J. Gomis, J. M. Pons, and N. Roman-Roy (1986). Equivalence between the Lagrangian and Hamiltonian formalism for constrained systems. *Journal of Mathematical Physics*, 27(12): 2953–2962.

Belot, Gordon (2003). Symmetry and gauge freedom. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, 34(2): 189–225.

Belot, Gordon (2007). The representation of time and change in mechanics. In Butterfield, J. and J. Earman, editors, *Handbook of the Philosophy of Physics Part A*.

Belot, Gordon (2013). Symmetry and equivalence. In Batterman, Robert, editor, *The Oxford Handbook of Philosophy of Physics*, pages 318–339. Oxford University Press.

Belot, Gordon (2018). Fifty million elvis fans can't be wrong. *Noûs*, 52(4): 946–981.

Belot, Gordon and John Earman (2001). Pre-socratic quantum gravity. In *Physics Meets Philosophy at the Planck Scale*, pages 213–255. Cambridge University Press.

Bergmann, Peter G (1961). Observables in general relativity. *Reviews of Modern Physics*, 33(4): 510.

Bradley, Clara and James Owen Weatherall (2020). On representational redundancy, surplus structure, and the hole argument. *Foundations of Physics*, 50(4): 270–293.

Butterfield, Jeremy (2006). On symplectic reduction in classical mechanics. In Butterfield, Jeremy and John Earman, editors, *The Handbook of Philosophy of Physics*, pages 1–131. North Holland.

Curiel, Erik (2014). Classical mechanics is lagrangian; it is not hamiltonian. *The British Journal for the Philosophy of Science*, 65(2): 269–321.

Dasgupta, Shamik (2016). Symmetry as an epistemic notion (twice over). *The British Journal for the Philosophy of Science*, 67(3): 837–878.

Dewar, Neil (2019). Sophistication about symmetries. *The British Journal for the Philosophy of Science*, 70(2): 485–521.

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

Dirac, Paul A. M. (1964). *Lectures on Quantum Mechanics*. Dover.

Dittrich, Bianca (2006). Partial and complete observables for canonical general relativity. *Classical and Quantum Gravity*, 23(22): 6155.

Dittrich, Bianca (2007). Partial and complete observables for hamiltonian constrained systems. *General Relativity and Gravitation*, 39(11): 1891–1927.

Earman, John (2002a). Thoroughly modern mctaggart: or, what mctaggart would have said if he had read the general theory of relativity. *Philosophers' Imprint*, 2(3): 1–28.

Earman, John (2002b). Tracking down gauge: An ode to the constrained hamiltonian formalism. In Brading, Katherine and Elena Castellani, editors, *Symmetries in Physics: Philosophical Reflections*, pages 140–62. Cambridge University Press.

Earman, John (2004). Laws, symmetry, and symmetry breaking: Invariance, conservation principles, and objectivity. *Philosophy of Science*, 71(5): 1227–1241.

Fletcher, Samuel C (2020). On representational capacities, with an application to general relativity. *Foundations of Physics*, 50(4): 228–249.

Gotay, Mark J and James M Nester (1979). Presymplectic lagrangian systems. i: the constraint algorithm and the equivalence theorem. In *Annales de l'institut Henri Poincaré. Section A, Physique Théorique*, volume 30, pages 129–142.

Gotay, Mark J and James M Nester (1980). Presymplectic lagrangian systems. ii: the second-order equation problem. In *Annales de l'institut Henri Poincaré. Section A, Physique Théorique*, volume 32, pages 1–13.

Gotay, Mark J and James M Nester (1984). Apartheid in the dirac theory of constraints. *Journal of Physics A: Mathematical and General*, 17(15): 3063.

Gracia, X and JM Pons (1988). Gauge generators, dirac's conjecture, and degrees of freedom for constrained systems. *Annals of Physics*, 187(2): 355–368.

Gryb, Sean and Karim Thébault (2016a). Regarding the 'hole argument'and the 'problem of time'. *Philosophy of Science*, 83(4): 563–584.

Gryb, Sean and Karim Thébault (2016b). Schrödinger evolution for the universe: Reparametrization. *Classical and Quantum Gravity*, 33(6): 065004.

Gryb, Sean and Karim Thébault (2016). Time remains. *The British Journal for the Philosophy of Science*, 67(3): 663–705.

Gryb, Sean and Karim Thébault (2023). *Time Regained: Volume 1: Symmetry and Evolution in Classical Mechanics*. Oxford University Press.

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

Halvorson, Hans (2016). Scientific theories. In Humphreys, Paul, editor, *The Oxford Handbook of Philosophy of Science*. Oxford University Press.

Halvorson, Hans and Dimitris Tsementzis (2017). Categories of scientific theories. In *Categories for the Working Philosopher*. Oxford University Press.

Henneaux, Marc and Claudio Teitelboim (1994). *Quantization of Gauge Systems*. Princeton University Press.

Ismael, Jenann and Bas C Van Fraassen (2003). Symmetry as a guide to superfluous theoretical structure. *Symmetries in physics: Philosophical reflections*, pages 371–392.

Kuchař, Karel (1991). The problem of time in canonical quantization. In Ashtekar, A. and J. Stachel, editors, *Conceptual Problems of Quantum Gravity*, pages 141–171. Birkhäuser.

Malament, David B. (2012). *Topics in the Foundations of General Relativity and Newtonian Gravitation Theory*. University of Chicago Press.

Martens, Niels C.M. and James Read (2021). Sophistry about symmetries? *Synthese*, 199(1-2): 315–344.

Maudlin, Tim (2002). Thoroughly muddled mctaggart: Or, how to abuse gauge freedom to create metaphysical monostrosities. *Philosopher's Imprint*, 2(4).

Nguyen, James, Nicholas J. Teh, and Laura Wells (2020). Why surplus structure is not superfluous. *The British Journal for the Philosophy of Science*, 71(2): 665–695.

North, Jill (2009). The “structure" of physics: A case study. *The Journal of Philosophy*, 106(2): 57–88.

Pitts, J. Brian (2014a). Change in hamiltonian general relativity from the lack of a time-like killing vector field. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, 47: 68–89.

Pitts, J. Brian (2014b). A first class constraint generates not a gauge transformation, but a bad physical change: The case of electromagnetism. *Annals of Physics*, 351: 382–406.

Pitts, J. Brian (2022). First-class constraints, gauge transformations, de-ockhamization, and triviality: Replies to critics, or, how (not) to get a gauge transformation from a second-class primary constraint. *https://philsci-archive.pitt.edu/21423/*.

Pitts, J. Brian (2024). Does a second-class primary constraint generate a gauge transformation? electromagnetisms and gravities, massless and massive. *Annals of Physics*, 462: 169621.

Pons, J.M., D.C. Salisbury, and L.C. Shepley (1999). Reduced phase space: quotienting procedure for gauge theories. *Journal of Physics A: Mathematical and General*, 32(2): 419.

Pons, Josep M. (1988). New relations between hamiltonian and lagrangian constraints. *Journal of Physics A: Mathematical and General*, 21(12): 2705.

Pons, Josep M. (2005). On dirac’s incomplete analysis of gauge transformations. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, 36(3): 491–518.

Pons, J. M., D. C. Salisbury, and L. C. Shepley (1997). Gauge transformations in the lagrangian and hamiltonian formalisms of generally covariant theories. *Phys. Rev. D*, 55: 658–668.

Pooley, Oliver and David Wallace (2022). First-class constraints generate gauge transformations in electromagnetism (reply to pitts). *arXiv preprint arXiv:2210.09063*.

Roberts, Bryan W (2020). Regarding ‘leibniz equivalence’. *Foundations of Physics*, 50: 250–269.

Rovelli, Carlo (2002). Partial observables. *Physical Review D*, 65(12): 124013.

Rovelli, Carlo (2004). *Quantum Gravity*. Cambridge University Press.

Rovelli, Carlo (2014). Why gauge. *Foundations of Physics*, 44: 91–104.

Souriau, Jean-Marie (1997). *Structure of dynamical systems: a symplectic view of physics*, volume 149. Springer Science & Business Media.

Thébault, Karim (2012). Symplectic reduction and the problem of time in nonrelativistic mechanics. *The British Journal for the Philosophy of Science*, 63(4): 789–824.

Thébault, Karim (2021). The problem of time. In Knox, Eleanor and Alastair Wilson, editors, *The Routledge Companion to Philosophy of Physics*. Routledge.

Thiemann, Thomas (2008). *Modern Canonical Quantum General Relativity*. Cambridge University Press.

Wallace, David (2002). Time-dependent symmetries: The link between gauge symmetries and indeterminism. In Brading, Katherine and Elena Castellani, editors, *Symmetries in Physics: Philosophical Reflections*, pages 163–173. Cambridge University Press.

Wallace, David (2022). Observability, redundancy, and modality for dynamical symmetry transformations. In Read, James and Nicholas J. Teh, editors, *The Philosophy and Physics of Noether's Theorems: A Centenary Volume*, page 322–353. Cambridge University Press.

Weatherall, James Owen (2016a). Are newtonian gravitation and geometrized newtonian gravitation theoretically equivalent? *Erkenntnis*, 81: 1073–1091.

Weatherall, James Owen (2016b). Understanding gauge. *Philosophy of Science*, 83(5): 1039–1049.

Weatherall, James Owen (2017). Category theory and the foundations of classical space–time theories. In *Categories for the Working Philosopher*. Oxford University Press.

Weatherall, James Owen (2018). Regarding the 'hole argument'. *The British Journal for the Philosophy of Science*, 69(2): 329–350.

Weatherall, James Owen (2019a). Part 1: Theoretical equivalence in physics. *Philosophy Compass*, 14(5): e12592.

Weatherall, James Owen (2019b). Part 2: Theoretical equivalence in physics. *Philosophy Compass*, 14(5): e12591.

Appendices

A Chapter 1

A.1 Proposition 1.2

We want to show that F is not full i.e. that it fails to be surjective on arrows, but that it is faithful and essentially surjective.

To show that F is not full, we need to show that there is an arrow $g : (\Sigma_f, \tilde{\omega}_f, H) \rightarrow (\Sigma_f, \tilde{\omega}'_f, H')$ such that $g \neq F(f)$ for any arrow f in **TotHam**. To do this, we will show that there are transformations along the vector fields associated with the secondary first-class constraints that are not arrows in **TotHam**, but their restriction to Σ_f are arrows in **ExtHam**. Consider the diffeomorphism $f : \Sigma_p \rightarrow \Sigma_p$ that takes each point on Σ_p to another point along the vector field associated with the secondary first-class constraints at that point. Recall that $\tilde{\omega}_p(X_{\varphi_j}, \cdot) = d\varphi_j \neq 0$ for the secondary first-class constraints φ_j, which tells one the change of a function along the vector fields associated with the secondary first-class constraints. So let us consider f to be the flow of the vector field associated with $\alpha^j d\varphi_j$ where α is an arbitrary function of the canonical coordinates. In order to associate a vector field X with $\alpha^j d\varphi_j$ via $\tilde{\omega}_p(X, \cdot) = \alpha^j d\varphi_j$, it must be that $\alpha^j d\varphi_j$ is exact i.e. can be represented as $d\gamma$ for some scalar field γ. Therefore, it must be that $d(\tilde{\omega}_p(X, \cdot)) = d(\alpha^j d\varphi_j) = 0$. But

$d_b(\alpha^j d_a \varphi_j) = \alpha^j d_b d_a \varphi_j + d_{[a}\varphi_j d_{b]}\alpha^j$. The first term vanishes since $d(d\varphi_j) = 0$ by Poincaré's Lemma. However, the second term does not necessarily vanish, since α^j is an arbitrary function of the canonical coordinates (so $d\alpha^j$ is not necessarily zero). In such cases, one cannot associate with f a vector field via $\tilde{\omega}_p(X, \cdot) = \alpha^j d\varphi_j$. This means that in these cases the flow of the vector field associated with $\alpha^j d\varphi_j$ does not consist of symplectomorphisms[17], and so $f^*(\tilde{\omega}_p) \neq \tilde{\omega}_p$. Therefore, f is not an arrow in **TotHam** (for every choice of α^j).

However, let us consider the transformations along the vector fields associated with the secondary first-class constraints on Σ_f. That is, consider the (gauge) transformation $g : \Sigma_f \to \Sigma_f$ that is a diffeomorphism that takes each point on Σ_f to another arbitrary point along the gauge orbit associated with the secondary first-class constraints φ_j at that point. Since $d\varphi_j = 0$ on the final constraint surface, $\alpha^j d\varphi_j = 0$ where α^j is an arbitrary function. But this means that one *can* associate a vector field Y with $\alpha^j d\varphi_j$ via $\tilde{\omega}_f(Y, \cdot) = \alpha^j d\varphi_j$ since $d(\alpha^j d\varphi_j) = 0$ on Σ_f . Therefore, the flow of Y consists of symplectomorphisms. Moreover, $g^*H = H$ because H is gauge-invariant on the final constraint surface. Therefore, g is an arrow in **ExtHam** (for all choices of α^j). This implies that there are arrows g of **ExtHam** such that $g \neq F(f)$ for any arrow f in **TotHam**. Therefore, we can conclude that F is not full.

That F is essentially surjective follows from the fact that every object of **ExtHam** is the restriction of some object $(\Sigma_p, \tilde{\omega}_p, H, \varphi_i)$ to the surface defined by $\varphi_i = 0$. Finally, to show that F is faithful, we need to show that if two arrows f, g between objects $(\Sigma_p, \tilde{\omega}_p, H, \varphi_i)$, $(\Sigma_p, \tilde{\omega}'_p, H', \varphi'_i)$ of **TotHam** are distinct, then their action on Σ_f is distinct. In other words, we want to show that if $f|_{\Sigma_f} = g|_{\Sigma_f}$, then $f = g$. So suppose that $f|_{\Sigma_f} = g|_{\Sigma_f}$. If f, g are arbitrary gauge transformations, then the only way that f and g could differ is if they move points off of Σ_f by differing amounts along the vector fields associated with the primary first-class constraints. But H is not constant along the vector fields associated with the

[17] See Abraham and Marsden (1987) Proposition 3.3.6.

primary first-class constraints off of Σ_f. Since f, g must preserve H by definition, f must be equal to g. If f, g are symplectomorphisms that are flows along vector fields other than the vector fields associated with the primary first-class constraints, then the only way that f and g could differ is if at least one changes the secondary constraints. But since f, g must preserve φ_i by definition, f must be equal to g. So F is faithful.

B Chapter 2

B.1 Proposition 2.1

Suppose that $\tilde{\omega}_p(X, \cdot) = 0$ i.e. X is a null vector field on Σ_p. This implies that $\tilde{\omega}_p(X, \cdot) \circ FL_p = FL_p^*(\tilde{\omega}_p(X, \cdot)) = 0$ since FL_p is a submersion. If X is of the form $FL_{p*}(Z)$ for Z on T_*Q, then we can write this as $FL_p^*(\tilde{\omega}_p(FL_{p*}(Z), \cdot)) = 0$, which is equivalent to $(FL_p^*\tilde{\omega}_p)(Z, \cdot) = 0$ by Malament (2012, Proposition 1.5.1). Since $FL_p^*\tilde{\omega}_p = \Omega$, this implies that Z is a null vector field of Ω. But every distinct vector field X on Σ_p can be written as $FL_{p*}(Z)$ for distinct vector fields Z on T_*Q. In particular, we think of Z as a vector field whose action at all points in the inverse image of some point in Σ_p, $FL_p^{-1}(FL_p(x))$, on smooth maps of the form $FL_p^*(f)$ is given by the action of X on f at $FL_p(x)$. This shows that every distinct null vector field on Σ_p corresponds to a distinct null vector field on T_*Q.

Next, we want to show that the only additional null vector fields of Ω are the vector fields in $Ker(FL_{p*})$. First, we will show that if Z is a null vector field of Ω, then when $FL_{p*}(Z)$ is well-defined, $FL_{p*}(Z)$ is a null vector field on Σ_p i.e. there does not exist a null vector field of Ω whose pushforward along FL_p defines a vector field on Σ_p that is not null with respect to $\tilde{\omega}_p$. Then, we will show that $Ker(FL_{p*}) \subseteq Ker(\Omega)$ i.e. the vector fields in $Ker(FL_{p*})$ are null vector fields of Ω.

For the first, suppose that $\Omega(Z, \cdot) = 0$. By the definition of Ω, this means that at all points $x \in T_*Q$, $(FL_p^*\tilde{\omega}_p)(Z, \cdot) = 0$. This is equivalent to $\tilde{\omega}_p(FL_{p*}(Z), \cdot) = 0$ at the point $FL_p(x)$ i.e. $FL_p^*(\tilde{\omega}_p(FL_{p*}(Z), \cdot)) = 0$ for all points $x \in T_*Q$. Since FL_p is a submersion, this means that $\tilde{\omega}_p(FL_{p*}(Z), \cdot) = 0$ at all points $FL_p(x) \in \Sigma_p$. This means that for points $x \in T_*Q$ such that $FL_{p*}(Z)$ is a well-defined vector field on Σ_p, $FL_{p*}(Z)$ is a null vector field of $\tilde{\omega}_p$.

For the second, suppose that $Y \in Ker(FL_{p*})$. Then, from the above, $\Omega(Y, \cdot) = (FL_p^*\tilde{\omega}_p)(Y, \cdot) = \tilde{\omega}_p(FL_{p*}(Y), \cdot)$. But $FL_{p*}(Y)$ is the zero vector at every point, and so $\Omega(Y, \cdot) = 0$. This means that $Ker(FL_{p*}) \subseteq Ker(\Omega)$ and so there are multiple distinct null vector fields on T_*Q that correspond to the trivial (zero) null vector field on Σ_p.

Therefore, the number of null vector fields on T_*Q is equal to the number of null vector fields on Σ_p plus the dimension of $Ker(FL_{p*})$, where the dimension of $Ker(FL_{p*})$ is equal to the number of primary first-class constraints.

B.2 Proposition 2.2

In order to use the same proof that was used for Proposition 2.1, we need to show that FL_f is a (surjective) submersion. To see why FL_f is a submersion, notice that Proposition 2.1 implies that if $dE(Z)$ is a Lagrangian constraint where Z is a null vector field on T_*Q, then $dH(FL_{p*}(Z))$ is a Hamiltonian constraint. Similarly, if $dH(X)$ is a Hamiltonian constraint where X is a null vector field on Σ_p, then $dE(Z)$ is a Lagrangian constraint where $X = FL_{p*}(Z)$. Moreover, $dE(Y)$ for $Y \in Ker(FL_{p*})$ is automatically zero, since by assumption of almost regularity E is constant along the fibers $FL^{-1}(FL(q, \dot{q}))$. This means that there will be a one to one correspondence between Lagrangian constraints of this kind and the first generation of secondary Hamiltonian (first-class) constraints. Reiterating, the same will be true of all further constraint submanifolds, and so since each constraint reduces the dimension by one, the relationship between P_f and Σ_f will be the same relationship as between T_*Q

and Σ_p: the induced Legendre transformation FL_f will be a surjective submersion, where $Ker(FL_{f*}) = Ker(FL_{p*})$.

Therefore, we can use the same proof as the proof for Proposition 2.1 to show that distinct null vector fields on Σ_f correspond to distinct null vector fields on P_f, and that there are additional null vector fields on P_f corresponding to $Ker(FL_{f*}) = Ker(FL_{p*})$. So the number of null vector fields on P_f is equal to the number of first-class constraints plus the dimension of $Ker(FL_{p*})$, which recall is the number of primary first-class constraints.

B.3 Proposition 2.3

Similar to the proof of Proposition 2.1, suppose that $\tilde{\omega}_f(X_H, \cdot) = dH$. This implies that $FL_f^*(\tilde{\omega}_f(X_H, \cdot)) = FL_f^*(dH)$ since FL_f is a submersion. If X_H is of the form $FL_{f*}(X_E)$ for some vector field X_E on P_f, then we can write this as $FL_f^*(\tilde{\omega}_f(FL_{f*}(X_E), \cdot)) = FL_f^*(dH)$, which is equivalent to $(FL_f^*\tilde{\omega}_f)(X_E, \cdot) = FL_f^*(dH)$. Since $FL_f^*\tilde{\omega}_f = \Omega_f$ and $FL_f^*(dH) = d(FL_f^*H) = dE$, this implies that X_E is a solution to $\Omega_f(X_E, \cdot) = dE$, which is the equations of motion on P_f. Since every (distinct) vector field X_H on Σ_f can be written as $FL_{f*}(X_E)$ for (distinct) vector fields X_E on P_f, this shows that every distinct solution on Σ_f corresponds to a distinct solution on P_f.

Now suppose that $\Omega_f(X_E, \cdot) = dE$. By the definition of Ω_f and E, this means that $(FL_f^*\tilde{\omega}_f)(X_E, \cdot) = FL_f^*(dH)$. This is equivalent to $\tilde{\omega}_f(FL_{f*}(X_E), \cdot) = FL_f^*(dH)$ at the point $FL_f(x)$ i.e. $FL_f^*(\tilde{\omega}_f(FL_{f*}(X_E), \cdot)) = FL_f^*(dH)$ for all points $x \in P_f$. Since FL_f is a submersion, this means that $\tilde{\omega}_f(FL_{f*}(X_E), \cdot) = dH$ for all points $FL_f(x) \in \Sigma_f$. This shows that when $FL_{f*}(X_E)$ is a well-defined vector field on Σ_f, $FL_{f*}(X_E)$ is a solution on Σ_f.

Finally, we want to show that the relationship between solutions is many to one. This follows from the fact that if X_E is a solution to $\Omega_f(X_E, \cdot) = dE$, then so is $X_E + \alpha^i Y_i$ where

$Y_i \in Ker(FL_{f*})$ and α^i is an arbitrary function on Ω_f since $Ker(FL_{f*}) \subseteq Ker(\Omega_f)$. But $FL_{f*}(X_E + \alpha^i Y_i) = FL_{f*}(X_E)$. Therefore, there are distinct solutions on P_f that correspond to the same solution on Σ_f.

B.4 Proposition 2.4

To show that J is a functor, we need to to show that J takes objects of **LagR** to objects of **HamR** and arrows to arrows. The first is trivial. To show the second, take an arrow f between objects $(\bar{P}_1, \bar{\Omega}_1, \bar{L}_1)$ and $(\bar{P}_2, \bar{\Omega}_2, \bar{L}_2)$. Since f is a symplectomorphism, $f^*\bar{\Omega}_2 = \bar{\Omega}_1$. Since $\bar{\Omega} = F\bar{L}^*\bar{\omega}$ by construction, this means that $f^*(F\bar{L_2}^*\bar{\omega}_2) = F\bar{L_1}^*\bar{\omega}_1$. We want to show that $F\bar{L}_2 \circ f \circ F\bar{L}_1^{-1}$ is an arrow in **HamR**. That is, we want to show that $(F\bar{L}_2 \circ f \circ F\bar{L}_1^{-1})^*\bar{\omega}_2 = \bar{\omega}_1$ and $(F\bar{L}_2 \circ f \circ F\bar{L}_1^{-1})^*(\bar{E}_2 \circ F\bar{L_2}^{-1}) = \bar{E}_1 \circ F\bar{L}_1^{-1}$. The first follows from the fact that $f^*(F\bar{L_2}^*\bar{\omega}_2) = F\bar{L_1}^*\bar{\omega}_1$. The second follows from the fact that $f^*\bar{E}_2 = \bar{E}_1$ since $f^*\bar{L}_2 = \bar{L}_1$. Similar reasoning can be used to show that K is a functor.

Since $F\bar{L}$ and $F\bar{H}$ are global diffeomorphisms, one can define the inverse $F\bar{L}^{-1} = F\bar{H}$ and $F\bar{H}^{-1} = F\bar{L}$. This implies that the functors J and K are inverses on objects and similarly on arrows. That J and K preserve solutions follows from the fact that FL_f preserves solutions (from Proposition 2.3), and that the solutions that are equivocated through reduction are just the gauge-related solutions.

C Chapter 4

C.1 Proposition 4.1

It is easy to show by example that there are gauge transformations that do not preserve the two-form $\omega = \sum_i dq^i \wedge dp_i$ in coordinate form. For example, consider the infinitesimal transformation $q_i' = q_i + \epsilon a^j(q_i,p_i)\{q_i,\varphi_j\}$ where we set $a^j(q_i,p_i)$ such that $p_i' = p_i$. In general, this will fail to preserve ω. In particular, it will fail to preserve ω when $d(a^j(q_i,p_i)\{q_i,\varphi_j\}) \neq 0$, which is just when $d(a^j(q_i,p_i)) \neq 0$ since $d\omega = 0$ (it is closed). Since all point*-transformations are symplectomorphisms, this means that arbitrary gauge transformations will also not be point*-transformations.

More generally, we can show that arbitrary gauge transformations, namely, diffeomorphisms $g : T^*Q \to T^*Q$ that act by moving points of cotangent space an arbitrary amount along the vector fields associated with the first-class constraints at each point, are such that $g^*\omega' \neq \omega$ where ω' is the two-form under the transformation generated by g. The proof has the same structure as the first part of the proof for Proposition 1.1, except we now do not need to restrict to *secondary* first-class constraints: Take the gauge transformations to be represented by the flow of the vector field associated with $\alpha^i d\varphi_i$ where α is an arbitrary function of the canonical coordinates and φ_i are the first-class constraints. Then because $\omega(X_{\varphi_i},\cdot) = d\varphi_i \neq 0$, it follows that one cannot associate with $\alpha^i d\varphi_i$ a vector field via $\omega(Y,\cdot) = \alpha^i d\varphi_i$ unless $d_{[a}\varphi_i d_{b]}\alpha^i = 0$. This means that the flow of the vector field associated with $\alpha^i d\varphi_i$ does not consist of symplectomorphisms for all gauge transformations g and so $g^*\omega' \neq \omega$ in general.

C.2 Proposition 4.2

This follows the same structure as the second part of the proof of Proposition 1.1: Take the gauge transformations on the final constraint surface to be represented by the flow of the vector field associated with $\alpha^i d\varphi_i$. Since $d\varphi_i = 0$ for all first-class constraints on the final constraint surface, $\alpha^i d\varphi_i = 0$. This means that one can associate a vector field Y with $\alpha^i d\varphi_i$ via $\tilde{\omega}_f(Y,\cdot) = \alpha^i d\varphi_i$ since $d(\alpha^i d\varphi_i) = 0$ on Σ_f . Therefore, the flow of Y consists of symplectomorphisms, and so gauge transformations are symplectomorphisms on the final constraint surface. Moreover, $g^*H = H$ because H is gauge-invariant on the constraint surface. Therefore, gauge transformations are arrows in **ConHam2** (in fact, this shows that they are automorphisms).

C.3 Proposition 4.3

Since reduction equivocates between points along the gauge orbits, this proposition follows trivially, since gauge transformations are just the identity transformation on the reduced space, which are arrows in **ConHamRed**.

C.4 Proposition 4.4

That F is essentially surjective follows from the fact that every model of the reduced theory is the reduction of some model on the final constraint surface by definition. That F is full follows from the fact that all arrows $\bar{g}$ between $F(\Sigma_f, \tilde{\omega}_f, H)$, $F(\Sigma'_f, \tilde{\omega}'_f, H')$ are arrows between $(\Sigma_f, \tilde{\omega}_f, H), (\Sigma'_f, \tilde{\omega}'_f, H')$ defined on the equivalence class of points along the gauge orbits.

That F is not faithful can be shown by demonstrating that there are two distinct arrows in **ConHam2** that map to the same arrow in **ConHamRed**. Consider some model $(\Sigma_f, \tilde{\omega}_f, H)$ where $\tilde{\omega}_f$ has at least one null vector field and consider two arrows g_1, g_2 from $(\Sigma_f, \tilde{\omega}_f, H)$ to itself corresponding to distinct gauge transformations i.e. two different ways of moving the points of Σ_f to other points along the gauge orbits (as long as $\tilde{\omega}_f$ has at least one null vector field, one can find such distinct g_1, g_2). Then $F(g_1) = F(g_2) = Id$ since g_1 and g_2 preserve the gauge orbits. So F is not faithful.