A Prescriptivist Account of Physical Theories

by

Shahin Kaveh

BSc in Physics, Sharif University of Technology, 2008

MA in Philosophy, Texas Tech University, 2011

MA in Philosophy, Johns Hopkins University, 2014

Submitted to the Graduate Faculty of

the Dietrich School of Arts and Sciences in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2021

UNIVERSITY OF PITTSBURGH

DIETRICH SCHOOL OF ARTS AND SCIENCES

This dissertation was presented

by

Shahin Kaveh

It was defended on

July 28th 2021

and approved by

John D. Norton, History and Philosophy of Science

John Earman, History and Philosophy of Science

James Woodward, History and Philosophy of Science

Richard Healey, Philosophy (Univeristy of Arizona)

A Prescriptivist Account of Physical Theories

Shahin Kaveh, PhD

University of Pittsburgh, 2021

A question of central importance to any philosopher of science is: what is the essential content of a scientific theory? What does a given theory really tell us about the world? Philosophers of science have disagreed on many aspects of the answer to this question, for instance whether the essential content of theories concerns entities, properties, or structures, whether it should be cashed out in terms of sentences or models, and whether one should be a realist or an anti-realist about this content; but philosophers have near-universally agreed on one claim: that theories provide a description of the natural system to which they are applied. Call this the descriptive-ontological view. I argue against the descriptive-ontological view in physics and propose an alternative: the prescriptive-dynamical view. According to the latter, the essential content of a physical theory is to provide prescriptions for interfacing with the natural system. More precisely, physical theories consist of a fixed part and an open-ended part, such that the fixed part is a prescription for constructing the open-ended part from local data, gathered through interaction with the system. The answer to the question of essential content directly determines or at least influences one's response to many other crucial questions such as theoretical equivalence (Chapter 2), theory-world relations (Chapter 3), and realism-antirealism (Chapter 4), which I will subsequently explore. Moreover, as I will argue (Chapter 5), the prescriptive-dynamical account also sheds fresh light on the history of quantum mechanics as a painstaking process of realization that instead of telling us what there is, physical theories must tell us what to do.

Table of Contents

List of Tables

List of Figures

Preface

In addition to John D. Norton, James Woodward, John Earman, and Richard Healey, I am grateful to Anjan Chakravartty, Kyle Stanford, Porter Williams, Gal Ben Porath, Dana Matthiessen, Aydin Mohseni, and numerous conference participants and anonymous reviewers for invaluable discussion and commentary.

1.0 Chapter 1: Physical Theories Are Prescriptions, Not Descriptions

1.1 Introduction

I argue that fundamental physical theories have a fixed part and an open-ended part, such that the fixed part is a *prescription for constructing the open-ended part from local empirical data*. Thus, the essential content[1] of physical theories is in a precise sense a prescription for how to interface with natural systems rather than a description of what (entity, property, or structure) there is. The descriptive aspects of theories might serve as embellishment or scaffolding, but they are inessential to the predictive success of the theory.

I will start in Section 1.2 by providing a brief exposition of the distinction between fixed and open-ended formulae, as well as a brief explanation of what is amiss in traditional accounts. Then, in Section 1.3, I go on to provide textual evidence that the two traditional accounts of theories (syntactic and semantic views) have assumed fundamental predictive theories to be *fixed* rather than open-ended. Section 1.4 is devoted to a more in-depth exploration of the characteristics of fixed and open-ended equations, and their distinct relations to the natural system and the observer. My considerations culminate in Sections 1.5 and 1.6, in which I formulate what I call the *prescriptive-dynamical view* (§1.5) and show how it likely rules out all forms of the *descriptive-ontological view* (§1.6), respectively.

1.2 Brief Exposition of the Fixed / Open-ended Distinction

1.2.1 The two-tier character of theories

The following is a brief overview of the distinction I have in mind (a more in-depth explanation to come in Section 1.4): upon examining the various fundamental formulae that

[1]The notion of essential content is akin to Kitcher's *working posits* (1993, 149 ff.) and Psillos's "indispensable" or "essential contributors" (1999, 108 ff.), which are characterized as those parts of the theory that actually contribute to its empirical success, as opposed to any idle baggage that might come with the theory. See §1.6.1 below for further discussion.

are used to make predictions about a natural system, such as when Newtonian mechanics is applied to a physical pendulum or the solar system, we find that *predictions are made using a formula that remains the same in every application of the theory plus a formula that is different from context to context and thus must be determined based on the specific interaction and the natural system being modelled.* For instance, $F = ma$ is the *fixed* principle in Newtonian mechanics, while a "force law" such as $F = -kx$, $F = -bv^2$, or $F = -\frac{Gmm'}{r^2}$ supplies the *open-ended* formula. It is only the combination of the two that yields empirical predictions, for their conjunction yields a differential equation that takes us from initial conditions to final conditions. The same pattern exists in every theory that allows for a dynamical systems formulation (see Table 1).

Table 1: The most canonical dynamical theories of physics have a two-tier character with one fixed formula and one open-ended formula.

Theory	**Firxed Formula**	**Open-ended formula**
Newtonian mechanics	Newton's second law	Various force formulae
Non-rel. quantum mechanics	Schrödinger equation	Various Hamiltonians
Quantum field theory	Canonical commutation eqs.	Interaction Lagrangians
Bohr's atomic theory	Bohr quantization condition	Correspondence Principle
Diffusion theory	Diffusion equation	Diffusion functions
Standard cosmology	Friedmann equations	Eqs. of state and matter
Electrostatics	Poisson's equation	Density functions

As I shall explain in more detail in Section 1.4, the second group of formulae are *open-ended* in the sense that there is no exhaustive list of them to choose from when one comes across a new natural system being studied for the first time, or when interacting with the same system in a new context. There is no principle from which these formulae are derived, nor any way (other than trial and error) to know under what physical circumstances each must remain valid. Rather, they are "reverse engineered" from the fixed formula and the experimental results together. The open-ended formulae are therefore not *tested* against

data; they are *derived* by "fitting" the fixed formula to the data.

As said, my account is particularly suitable for dynamical systems. Any theory that features coherent state assignments and transition rules would be a dynamical system. In particular, a differential or integral equation can typically be read as expressing the evolution of a particular form of states, and thus theories governed by differential / integral equations are typically dynamical systems. Since virtually all theories of physics are dynamical systems, I expect my account to apply to physical theories broadly. Moreover, since the features I will be drawing on are quite generic among dynamical theories, one might reasonably expect my account to apply to other dynamical theories outside of physics, although again I will not be arguing for this general claim here. At the very least, I take this paper to establish the following claim: the most canonical theories at the foundation of contemporary physics exhibit a two-tier, open-ended character.[2]

This should be contrasted with the state of my competition, namely the syntactic and semantic construals of theories, which have had notorious difficulty in producing formulations of most scientific theories in accordance with their framework: how many predictive theories have we managed to couch as first-order axiomatic systems, and which theory have we managed to specify through an exhaustive list of its "class of models"? In light of this,

[2]Some constructs that may be classified as "theories", such as special relativity and statistical mechanics, are not dynamical theories and thus do not fall under the two-tier analysis. However, this is not a problem for my account, which concerns *predictive* theories. Statistical mechanics is a mathematical tool for bridging micro- and macro-theories, and any prediction that may be derived from it must originate in the dynamics of the micro-theory, rather than the posits of statistical mechanics itself. The case of special relativity is more complicated, but the sketch of a response can be provided as follows. Special relativity may be considered as a set of geometrical *coordination rules*, i.e. Lorentz transformations (which can still be taken as prescriptive though not open-ended) for aligning measuring rods and clocks in a world in which the maximum possible speed is some constant c. This theory makes no predictions of its own unless c is specified. Now, the statement that c is finite and equals the speed of light may be seen as external to the coordination rules of special relativity, either as a directly-verifiable empirical fact or as derived from Maxwell's dynamical equations. And as said, the light postulate is the sole source of predictive power in special relativity, to the extent that such power exists. Therefore, statistical mechanics and special relativity are not autonomous theories and yield no predictions of their own unless supplemented with dynamical theories. As such, they do not undermine my account.

Note also that sometimes a theory has both a dynamical formulation (with states evolving into other similar states) and a non-dynamical formulation. For example, General Relativity is sometimes presented as a yes / no function that takes a specified spacetime and matter distribution and tells us whether that combination is a physically possible universe. That is not a dynamical theory. But GR can also be formulated as a constrained Hamiltonian system, which takes states of the form (g^{ij}, π^{ij}) to other states of the same form, where g^{ij} is the 3-metric and π^{ij} its conjugate momentum (Lagrangian mechanics with its variational and dynamical formulations is another example).

I consider a table of the most foundational theories to which my account applies to be a promising start.

1.2.2 The problem with traditional accounts: being predictive while fixed

Nearly all philosophers of science speak of fundamental physical theories as fixed sets of sentences and/or models, in the sense that they take the theory to be the same set of sentences and/or models *across all local applications of the theory* (not to be confused with fixity across time – see below). There are a few notable exceptions from the modeling literature, most prominent among them Ronald Giere, who have recognized the existence of a fixed and a movable piece in fundamental theories.[3] Nevertheless, even these authors have failed to recognize the fixed piece as a prescription for constructing the movable piece through interaction with the system and the movable piece as truly local and open-ended (i.e. constrained by nothing but local data). Rather, these authors speak of the movable part as a simple addition to the fixed part that is chosen from a "family of models", one that is somehow constructed independently, and often based on a set of global / theoretical restrictions (rather than local / empirical facts).

The fixed part can perhaps be a set of axioms or class of models, but it is not predictive on its own. The entire theory which can make predictions, on the other hand, is not a fixed set of axioms or class of models. Thus, the problem arises when we use the phrase "the theory" to refer to something that has two attributes at the same time: i) it is fixed, i.e. remains the same across all successful applications, and ii) it is predictive, i.e. allows for empirical predictions in a robust range of contexts (when supplemented with auxiliaries and such). I shall argue in Section 1.4 that one finds nothing in the practice of science that has both attributes. But first allow me to provide textual evidence that simultaneous endorsement of i) fixity and ii) predictiveness is indeed widespread among philosophers of science.

[3]Cartwright has argued for the malleability of *phenomenological laws* but she, too, considers *fundamental laws* to be fixed. My claim is precisely that the latter are partially open-ended. More on this below.

1.3 Traditional Accounts of Theoretical Content

There are two main traditional accounts of what scientific theories are: the *syntactic view* and the *semantic view*. My objections to both accounts are the same, and the rivalry between them is independent of my dispute with both of them. However, since philosophers of science have mostly couched their views in terms of one or the other, let us deal with them in turn.[4]

1.3.1 Evidence from syntactic accounts

The theory in the syntactic view is usually treated as a partially interpreted, logically closed set of axioms whose theorems include observation statements derived with the help of initial conditions, auxiliaries, and correspondence rules.[5] These observation statements are then tested against empirical data.

A striking classic example is Hempel. In *The Theoretician's Dilemma*, Hempel describes a theory "as a set of sentences ... given in the form of axiomatized systems" (1958, 46). This supports the idea that the theory is fixed across successful applications: after all, if a theory is a set of sentences, then it would make no sense to say that the sentences of the theory are different from application to application: a different set of sentences would simply be a different theory.

> The laws referred to [here], such as the laws of Newtonian mechanics, are what we will call *statements of strictly universal form*, or *strictly universal statements*. (Hempel 1958, 39)

As said above, the problem is that Hempel also wants these fixed axioms to be predictive:

> The principles of Newtonian mechanics, for example, make it possible, given the present positions and momenta of the celestial objects that make up the solar system, to predict their positions and momenta for a specified future time or to postdict them for a specified time in the past. (Hempel 1958, 37)

[4]There is much debate about the equivalence or lack thereof of the syntactic and semantic views as well as their relative advantages. These debates will not affect my arguments below. See van Fraassen 1980; Lloyd 1988/1994; Suppe 2000; Halverson 2012; 2013; 2016.

[5]I will use "auxiliaries" to refer to background facts and theories of instrument, and "correspondence rules" to broadly mean any set of principles that connect theoretical terms to observable and/or measurable quantities (Halvorson 2016, 6).

Thus, Hempel explicitly endorses both (i) fixity and (ii) predictiveness.[6]

To be clear, I shall not criticize Hempel for claiming that $F = ma$ is a universal, exceptionless proposition. As is well-known, Cartwright has raised objections of this kind on the ground that the statements featuring in such predictive schemata are not strict generalizations and require ceteris paribus clauses (1983, 45 ff.; 1999, 24 ff.). My arguments below largely bracket the role of ceteris paribus clauses, if any. My point is rather that the fixed formulae of Newtonian mechanics are not predictive on their own, even if we were to add ceteris paribus clauses to them, and that they only become predictive when supplemented with a formula that is *open-ended*, which again may or may not be qualified by c-p clauses.

In Cartwright's terminology, my claim is that the *fundamental theoretical laws* (such as those from Table 1) are partially open-ended and local – something Cartwright herself strongly denies. She claims that *phenomenological laws* (e.g. amplifier models, exponential decay models, etc.) "tell what happens in concrete situations" whereas fundamental laws are "thoroughly abstract formulae which describe no particular circumstances." (Cartwright 1983, 11) Importantly, Cartwright's fundamental / phenomenological dichotomy cannot be mapped onto my fixed / open-ended distinction, because unlike fixed and open-ended formulae which must be *conjoined* for prediction to be possible, fundamental and phenomenological laws do not complement each other. Rather, they are said to work independently and at cross-purposes: a phenomenological law is said to single-handedly describe and predict the behavior of the system, but not explain it. A fundamental law, on the other hand, is taken as sufficient for explaining the behavior of the same system, but not for predicting it (ibid,

[6]Hempel proposes the following schema for such predictive inferences:

> C_1, C_2, ..., C_k
> L_1, L_2, ..., L_r
> ———————
> E
> Here, C_1, C_2, C_k are statements of particular occurrences (e.g., of the position and momenta of certain celestial bodies at a specified time), and L_1, ..., L_r are general laws (e.g., those of Newtonian mechanics); finally, E is a sentence stating whatever is being explained, predicted, or postdicted. (Hempel 1958, 37-8)

Both assumptions are clear in this passage: the theory is i) a set of "general laws" that ii) leads to empirical predictions in conjunction with "particular occurrences" such as initial conditions.

p. 3). The examples Cartwright gives for "fundamental laws" include "the equation of continuity and Boltzmann's equation" (ibid, p. 11), which suggest that Cartwright would put all formulae in Table 1 in the category of "fundamental laws", which she takes as fixed and global.

It is also important to distinguish my notion of fixity *across applications* from fixity *across time*, as is found for example in Lakatos's (1978) evolving research programs and Friedman's (2001) relativized a priori. Friedman argues that theories consist of an "a priori" part and an "a posteriori" part. The former can change occasionally when a "paradigm shift" occurs, but it is otherwise not directly susceptible to empirical modification, for it is needed for any empirical data to be possible at all – similar to Kant's transcendental conditions for the possibility of experience. The "a posteriori" part, however, consists of ordinary theories which respond directly to empirical data. As such, unlike my fixed formulae, Friedman's a priori principles are not prescriptions for local construction of the malleable part; they are global conditions for the possibility of any empirical knowledge at all, and thus transcend specific theories such as those in Table 1 and instead concern fundamental matters of space and time. By contrast, what I consider the fixed part of a theory here is typically a differential equation that prescribes rules for assigning dynamical states to the system. One might say that my entire discussion is located within the "a posteriori" section of Friedman's framework.

Regarding Lakatos, he posits that a research program consists of certain "core assumptions" that resist modification in light of new data, along with a "protective belt", i.e. a set of malleable assumptions which are allowed to adjust over time in order to protect "the hard core". While this framework can be useful in discussing the evolution of scientific frameworks over time, it still presents a theory *at any given time* as a set of global propositions that are *fixed for all applications*. To see the difference, consider the application of Newtonian mechanics to a physical mass-spring system. One often begins by using the secondary formula $F = -kx$ (somewhat unfortunately called Hooke's "law") for the restorative force of the spring. But this formula loses validity if the spring is stretched beyond a certain limit (which is different for different springs). At that point, terms such as x^2, $\dot{x}$, etc. might show up in the restorative force. Stretching the spring still further would continue to change this formula in ways that are hard to anticipate, and so would immersing it in a new medium,

heating it, subjecting it to stress or vibration, applying an external electromagnetic field, and so forth. For any new context, the secondary formula must be constructed anew from empirical data within the range of variables that characterize the local context.[7]

To force this into Lakatos's system, one would have to say that "the theory" (the conjunction of core and belt assumptions) is being "falsified" and revised every time the spring is further stretched, stressed, etc. This strikes me as an absurd and awkward characterization of what is going on. After all, the physicist never *assumes* that the most recently constructed force formulae are going to be universally valid, nor is she ever surprised or dismayed that the formula quickly loses relevance under new circumstances. On the contrary, that is precisely what she *expects* to happen. Hooke's "law" is not truly a law, and cannot be "falsified", for there is no prior presumption as to its exact form or scope of validity; it is simply an open-ended, local construction that depends on the context.[8]

Fixed characterizations of fundamental theories are by no means obsolete among the proponents of the syntactic view. Take for example Halvorson, who has recently mounted a vigorous defence of the syntactic view against the "bandwagon" of semantic conceptions (2012; 2013; 2016). He characterizes the syntactic view as follows:

> A scientific theory consists of two things:
> 1. A formal system, including:
> (a) Symbols;
> (b) Formation rules; and
> (c) Deduction rules.
> 2. Some use of this formal system to make claims about the physical world [using correspondence rules], and in particular, empirically ascertainable claims. (Halvorson 2016, 3)

Once again, a fixed set of axioms is taken to entail empirical predictions when conjoined with other auxiliaries. Indeed, the notion of a theory as a fixed set of axioms is taken so seriously that *any* set of axioms, including the empty set, is considered a theory (Halvorson 2012, 191).[9]

[7]That is, unless one has a way of *calculating* the secondary formula from microscopic equations that govern the particles in the spring, but that is more wishful thinking than anything related to the actual practice of physics.

[8]This is not to deny that such accounts can be useful in discussing theoretical change over time.

[9]Halvorson does criticize both syntactic and semantic views for taking theories to be "flat" rather than "structured". However, what Halvorson means by "structure" here is inferential hierarchy, which involves commitments to what is a logical consequence of what (including what is analytic and what is synthetic),

The above should suffice for demonstrating that syntactic accounts take theories to be both fixed and predictive.

1.3.2 Evidence from semantic accounts

On the semantic side, the class of models is usually picked out by a set of axioms or simply the intention of the scientist which delineates the class, and then structurally compared to empirical data through notions of isomorphism, homeomorphism, or sometimes broader notions of "similarity". The adherents of the semantic conception also take theories to be i) fixed classes or categories of models that ii) allow for empirical predictions in a range of contexts.

For a classic source, consider van Fraassen channelling Suppes[10]:

> Suppes's idea was simple: *to present a theory, we define the class of its models directly*, without paying any attention to questions of axiomatizabilizy, in any special language... . [I]f theories are to be reified – then a theory should be identified with its class of models. (van Fraassen 1989, 222)

That there is no open-ended component to theories in this view is fairly clear from the language of "class of models". Classes are defined through their membership and thus do not allow for open-endedness by definition.

But in case there are any lingering doubts about the fixity of theories in van Fraassen's view, consider the following. First of all, note that van Fraassen is simultaneously championing both the semantic view and the dynamical systems approach. In the dynamical approach, the essential content of a theory is given by two things: state assignments and transition rules. Hamiltonian mechanics, for instance, assigns states of the form (x, p) and encodes the transition rules among states of this form (permissible trajectories in the state space) in the quantity H. Once the state space and the transition rules are provided, the model is complete and can be used to make predictions. Now, it is clear that for van Fraassen, *both the state space and the transition rules are pre-determined* by the "class of models".

and so on (Halvorson 2012; 2016). But Halvorson's "structured" account still takes theories to be fixed.

[10]Much of van Fraassen and others' program was built on the pioneering work of Evert Willem Beth (e.g. 1960; see also: van Fraassen 1970).

> This introduces the idea of a cluster of models united by a common state-space; each has in addition a domain of objects plus a "history function" which assigns to each object a history, i.e. a trajectory in that space. (van Fraassen 1989, 223)

As we shall see below, the "history function", which provides the transition rules determining allowable trajectories in state space, is precisely the component that is open-ended. By assuming that the models are already restricted by a particular set of history functions, van Fraassen is implying that both the states and the transition rules are pre-ordained and hence fixed by the theory.

The same practice of treating both state assignments and transition rules as fixed and pre-ordained is also found in Frederick Suppe's work, who declares that for a realist semanticist: "a theory is empirically true just in case theory structure-allowed state transitions are identical to those possibly occurring in the actual world." (Suppe 2000, S106) Both states and transition rules are taken as pre-determined and tested against empirical data.

This brings us to claim (ii), namely that the fixed class of models is taken by semanticists to have predictive content. Van Fraassen explains:

> In Ronald Giere's recent encapsulation of the semantic approach, a theory consists of (a) the *theoretical definition*, which defines a certain class of systems; (b) a *theoretical hypothesis*, which asserts that certain (sorts of) real systems [such as the solar system] are among (or related in some way to) members of that class. (van Fraassen 1989, 222)

To be fair, Giere himself does not use the language of "classes" but "families", which might be taken as more loosely defined (1988; 2010). In fact, he is to my knowledge the only semantic theorist who has highlighted the "hierarchical" nature of physical theories in a manner similar to my "two-tier account" (see §1.2.1). He takes master formulae such as $F = ma$ to be "definitions" that allow physicists to construct more specific, "representational" models. The latter include specific secondary formulae such as force laws. This leads Giere to conclude that theories are not "well-defined entities", for there are "no necessary and sufficient conditions for what constitutes an admissible force function" (1988, 86 ff.). One might construe this as very close to saying that the theory is open-ended.

Not quite. In the account proposed here, *there are very clear necessary and sufficient conditions for what secondary formulae are allowed: namely that they be fitted to the local data under the guidance of the fixed formula.* The "open-endedness" of the secondary for-

mulae here refers not to the fact that there are no necessary and sufficient conditions, but that the necessary and sufficient conditions require conformity *only* to the local empirical data. Giere, on the other hand, assumes that scientists themselves provide the secondary formulae:

> [T]o be part of the theory of classical mechanics a model must bear a "family resemblance" to some family of models already in the theory. ... That question, it seems, is solely a matter to be decided by the judgments of members of the scientific community at the time. (Giere 1988, 86)

That is, Giere's secondary formulae are partially *pre-determined* by theorists and tested against data models (ibid; see also 2010, 271). Apart from the "collective judgment" of the scientific community, Giere says very little about how the movable pieces are constructed. This obscures both the prescriptive nature of the fixed part *and* the open-ended nature of the movable part. I shall argue below that such "family resemblance" restrictions only hamstring a dynamical theory.[11]

Once again, the approach outlined above is far from obsolete. For a more contemporary counterpart, consider Thomas Barrett's characterization of theories in his recent discussion of theoretical equivalence:

> *Criterion.* Theories T_1 and T_2 are equivalent according to the *model isomorphism* criterion if for every model of T_1 there is an isomorphic model of T_2, and vice versa. (Barrett 2017, 3)[12]

Isomorphism, as a bijective relationship, requires fixed sets. Furthermore, it is clear that

[11] While it is true that Giere talks about theories as "rules" for constructing concrete representational models, and that might sounds quite similar to "prescriptions", this talk for Giere is not specific to the fixed part of the theory: he considers the entire conjunction of the fixed and movable part as a set of rules for creating fully specified models of particular systems.

> The rules instruct one to locate the relevant masses and forces, and then to equate the product of the mass and acceleration of each body with the force impressed upon it. With luck one can solve the resulting equations of motion... . (Giere 1999, 94-5)

Giere's "rules" clearly encompass Newton's laws of motion *plus* a specific force function. My "prescriptions", by contrast, explain precisely how the latter is constructed from the combination of the former and local data in the first place.

[12] Halvorson makes this more precise in what he calls "pointwise isomorphism of models": "$\mathcal{M}$ is the same theory as $\mathcal{M}'$, just in case there is a bijection $F : \mathcal{M} \to \mathcal{M}'$ such that each model $m \in \mathcal{M}$ is isomorphic to its paired model $F(m) \in \mathcal{M}'$. (Halvorson 2012, 190)

Barrett considers these fixed sets of models to have "empirical content", which Barrett identifies with "the allowable trajectories for particles in the system" (ibid, 13).[13]

Thus, traditional accounts of science have ignored the open-endedness of predictive theories. As a note of reconciliation, a syntactic or semantic framework can still be useful in analyzing physical theories in certain contexts in which the locality and open-endedness of the formulae is not of concern to us (e.g. when remaining restricted to one singular application of the theory or when discussing a hypothetical "final theory"). But my point here is that these conditions are typically not met in the ordinary practice of physics.

To see how this leads to prescriptivism at the expense of descriptivism, we will need some preliminaries on the nature of open-endedness.

1.4 Preliminaries on the Nature of Open-ended Formulae

1.4.1 Local Empirical Mediating Principles (LEMPs)

Consider Newton's theory of mechanics. How is this theory applied in order to make predictions? To begin, the master formula $F = ma$ by itself does not provide any empirical predictions whatsoever, even if one were to add supplementary assumptions such as "correspondence rules" that connect F, m, and a to observables, as well as auxiliaries concerning theories of instrument, initial and boundary conditions, and ceteris paribus clauses. The fundamental reason for this is that F is not independently specified by Newton's theory. Thus, one needs a second formula that independently specifies F, something like $F = -kx$, $F = bv^n$, or $F = \frac{C}{r^4}$, say. Once this second formula is supplied, one can write down a truly predictive equation by eliminating F from the two relations. If the force formula is $F = -kx$, for instance, then the predictive relation is the differential equation $\ddot{x} = -\frac{k}{m}x$. With sufficient supplementary assumptions, this formula can be used to make predictions. Specifically, one needs two inputs, namely $(x, \dot{x})$, to be provided as initial conditions. Once

[13]There have been important recent developments in the modeling literature that add complexity to the traditional semantic accounts regarding the relationship between theory, model, and experiment (see: Morgan and Morrison 1999). However, all such nuanced views still fail to note the open-ended character of theories in the sense defended here. See, e.g., ibid, 3-4, for an endorsement of theoretically fixed models.

$(x(t_0), \dot{x}(t_0))$ are provided for some time $t = t_0$, the differential equation generates predictions for the values of $(x(t), \dot{x}(t))$ for all other t. This input-output dynamic is the core driver of predictive success in Newtonian mechanics.

This is of course true of all canonical theories of physics. In Lagrangian mechanics, the Lagrangian is the quantity that is not independently specified. In Schrödinger's quantum mechanics, it is the potential energy operator $\hat{V}$. In effective QFT, it is the interaction terms. And so on. All of these theories have a master formula that contains a quantity which is not independently specified, and which, once specified, together with the master formula would yield a predictive differential equation that takes us from initial conditions to final conditions. For reasons to be explained shortly, I shall call these secondary formulae *Local Empirical Mediating Principles* or LEMPs for short.

1.4.2 LEMPs are transition rules

The dynamical systems approach, which has seen a resurgence of interest in recent years[14], provides an excellent framework in which to easily understand the difference of content between fixed formulae and LEMPs. As said above, the dynamical approach involves focusing on states and transition rules as the essential content rather than on entities, properties, or structures. If one focuses on ontology, fixed formulae and LEMPs seem on a par: both attribute properties to objects. But from a dynamical standpoint, the difference of content is crystal-clear: the fixed formulae are always *state assignments*, while the LEMPs are always *transition rules*.

Define the *dynamical state* of a system as follows:

> The *dynamical state* of a system is the collection of measurable quantities of interest whose values at one time and place *would* uniquely determine their values at other times and places *were the system to be deterministic*.

In Newton's theory of mechanics, for example, the master formula $F = ma$ calls upon us to form differential equations of second order in position: $\ddot{x} = f(x, \dot{x}, t, ...)$, called *equations of motion*. This is equivalent to assigning dynamical states of the form $(x, \dot{x})$ to the system:

[14] See e.g. Wallace 2012; Rosaler 2015a; 2015b; Yoshimi 2012; Giunti 2006; 2016; Zednik 2011; Kaplan 2015; Ross 2015; Meyer 2018.

whenever the equation of motion has unique solutions, its solution is determined by a pair of inputs x and $\dot{x}$. Now, given initial values $(x(t_0), \dot{x}(t_0))$, the form of $f(x, \dot{x}, t, ...)$ will determine whether the future $(x(t), \dot{x}(t))$ will be unique and what values they can acquire. Thus, $f(x, \dot{x}, t, ...)$ encodes the transition rules. Since $f(x, \dot{x}, t, ...)$ is left unspecified, one can say that the fixed part of Newton's theory assigns states of the form $(x, \dot{x})$ to each degree of freedom in the system, while the open-ended force formulae such as $F \propto x$ and $F \propto \frac{1}{r^2}$ provide the transition rules among states of that form. The same goes for any other theory that can be given a dynamical systems formulation. If I am right, traditional accounts have gone wrong in assuming both the state assignments *and* the transition rules to be fixed by theory (see §1.3 above). In light of this, we may revisit the table of theories from Section 1.2 (see Table 2).

Table 2: The fixed formula in a dynamical theory provides the general form of state assignments while the open-ended formula encodes the missing ingredient for determining the transition rules among said states.

Theory	**State assignments (fixed)**	**Transition rules (open-end)**
Classical mechanics	$(x, \dot{x})$	F, L, or H
Non-rel. QM	experimental Hilbert states	evolution operator
Quantum field theory	Beams of sharp momentum $\|k\rangle$	S-matrix amplitudes
Bohr's atomic theory	Bohr's stationary states	Amplitudes permitted by CP
Diffusion theory	$(\rho(0, x), \rho(t, 0), \rho(t, R))$	diffusion function
Standard cosmology	(a, ρ)	equations of state and densities
Electrostatics	Dirichlet/Neumann boundaries	density function

Next I will explain in what sense the LEMPs must be considered open-ended.

1.4.3 Two aspects of open-endedness

In the next two subsections, I will elaborate on those characteristics of LEMPs which earn them the label "open-ended". At the beginning of this paper, I expressed the open-endedness

of LEMPs by saying that these formulae do not get *tested*; they are rather *derived* by "fitting" the master formula to the local empirical data in the first place. In the following, I will unpack this in terms of two notions of *abundance* (a fact about theories) and *inexhaustibility* (a fact about the world), which will prove important for my argument in Section 1.6.

1.4.3.1 Characteristic 1: LEMPs are abundant

There is only one condition for the existence of a predictively accurate LEMP: namely that the master formula be applicable to the system at hand. This is because the theory tasks us to *presuppose both the master formula and the experimental results* in a first encounter with a system and "reverse engineer" the LEMP from the two presuppositions such that the master formula plus the LEMP entail the experimental results. For example, suppose the natural system being studied is such that an acceleration $\ddot{x}$ can be assinged to it. Then it is not surprising that $\ddot{x}$ has *some* functional dependency on x and t: as the system moves through space, it will have *some* acceleration for any given position and time: $\ddot{x} = f(x, \dot{x}, t, ...)$. Therefore, whatever the observations are, one would simply assign the value of force to be whatever is needed to make $F = ma$ true, namely let $F = mf(x, \dot{x}, t, ...)$.[15]

The dynamical framework allows us to generalize this. Since fixed formulae are (the general form of) state assignment rules and LEMPs are (the missing ingredient for) transition rules, one could generally say that as long as the state assignments of the theory are applicable to the system at hand, there exist transition rules that predict the behavior of the system. For instance, as long as a system of dilute metallic gas gets deflected in a Stern-Gerlach machine, it can be assigned (detected and prepared in) various spin states. One can therefore record the state of the system as $s(t_1)$, $s(t_2)$, etc. at different times. One can then construct an evolution operator $D_{\delta t} : \{s(t_i)\} \to \{s(t_i)\}$ by demanding that $D_{t_2-t_1} s(t_1) = s(t_2)$, and so on for all recorded times. Whatever Hamiltonian function is the generator of this evolution operator will then give the Hamiltonian of the system. More generally we have:

> ABUNDANCE: The only condition for the existence of a predictively accurate LEMP is that the state assignments of the theory be applicable to the system at hand.[16]

[15] This is of course assuming that one has not previously constructed a suitable LEMP for the same system and the same context of interaction such that one already expects $f(x, \dot{x}, t, ...)$ to be a certain way.)

[16] Since LEMPs are abundant, the only reason to abandon a theory in this framework is if the state

One might object to ABUNDANCE by arguing that theories do put restrictions on LEMPs. For one thing, some transition rules may not be expressible as specific types of functions (e.g. Hamiltonian) due to mathematical restrictions on such functions. Moreover, one might impose additional requirements on LEMPs, such as requring that forces be deterministic, compliant with Newton's third law, conservative, time-independent, functions of position only, attributable to an identifiable source entity, and so on. As soon as one adds any such constraints beyond $F = ma$ itself, it becomes possible that one fails to find a suitable force equation that satisfies both conditions. For example, since Mercury can be assigned states of the form $(x, \dot{x})$, according to ABUNDANCE there is a force formula that predicts Mercury's orbit in conjunction with $F = ma$. One can confirm this by letting the force be $F = -Gmm'(\frac{1}{r^2} + \frac{3L^2}{r^4})$, where L is the angular momentum of the planet. This force predicts the anomalous precession of the perihelion of Mercury to the same accuracy as General Relativity.[17] However, this force explicitly depends on angular momentum. Thus, the objection goes, Newtonian theory fails to accommodate Mercury's orbit if one requires that gravity depend on relative distance only.

My response to this objection is that in practice, these additional constraints are easily discarded in favor of predictive power, and therefore they cannot be part of the essential

assignments of the theory become inapplicable to the system at hand. A thorough discussion of this condition is beyond the scope of this paper, but an illustration can be provided using the case of quantum mechanics.

One might begin by asking: if LEMPs are abundant, could we not find a modified force formula that would allow $F = ma$ to fit quantum phenomena? The answer is of course yes, we can. That is precisely what Bohmian quantum mechanics accomplishes. In Bohmian mechanics, one adds a "quantum potential" Q to the classical potential energy function V, and calculates the force as the gradient of this augmented potential function. The addition of Q accounts for all strange quantum behavior, evidenced by the fact that Bohmians can draw continuous trajectories for particles in a double-slit experiment.

So, then we ask: why do physicist consider quantum phenomena a reason to abandon $F = ma$ rather than modify the force formula? The prescriptivist answer would be that $F = ma$ has lost operational relevance: the state assignments of classical physics (namely x and $\dot{x}$) are still theoretically possible but no longer operationally assignable to quantum systems. This is because the correlations baked into the quantum potential Q make it impossible to prepare or detect the system in an arbitrary $(x, \dot{x})$ pair: attempting to prepare or detect the system in a particular x causes one to lose control over $\dot{x}$ and vice versa. So, while $F = ma$ is not technically "refuted" by quantum phenomena, it has lost its experimental relevance. This is why many practicing physicists state that Bohmian mechanics has no empirical relevance – a statement that makes perfect sense from a prescriptivist standpoint but is difficult to understand descriptively, given that Bohmian mechanics captures all of the same empirical data as its standard alternative.

In short, physicists switch to a new theory when the old state assignment rules become operationally inapplicable to the system. This can happen despite the fact that a modified LEMP can still be found for the theory to "save the phenomena".

[17] For a derivation, see Brown [manuscript].

content of the theory. As said above, as long as the state assignments are applicable, with enough patience one can always construct an adequate transition operator from the local data, and use the latter to form a predictive differential equation. There is no reason for the scientist to hamstring themselves and settle for less predictive power by ruling out certain transition rules in *a priori* fashion. For instance, Lagrangian and Hamiltonian mechanics are taken by the physicist as assigning states of the same form but encoding the transition rules in different quantities.[18] As such, physicists are quite opportunistic with respect to the different ways of encoding transition rules: in situations where mathematical restrictions prevent a certain type of function (e.g. Hamiltonian) from encoding certain empirical transition rules, the physicist swiftly moves to a different "formulation" of the theory (e.g. Lagrangian, etc.) in which the transition rules can be encoded. The Lagrangian formalism, for instance, is primarily used for unknown constraints and holonomic systems (see e.g. Goldstein 1980 [1950], 49), while Hamiltonian mechanics is mainly useful for systems with cyclic coordinates (see e.g. ibid, 351) and numerical calculations (see e.g. Finn 2008, 130). Even further, if the occasion calls for it, physicists will not shy away from mixing and matching Lagrangians and Hamiltonians in what they refer to as "Routhian" functions (Goldstein 1980 [1950], 352; Finn 2008, 137). Forces are used when the other methods do not pay off and/or are inapplicable (Goldstein 1980 [1950], 49, 351; Finn 2008, 130).

Or consider determinism. As Norton (2003; 2008) has argued, requiring that all forces be deterministic (causal fundamentalism) would preclude Newtonian mechanics from modeling many ordinary natural systems. Take the (in)famous example of a mass on top of a dome whose shape gives rise to the LEMP: $F = r^{\frac{1}{2}}$. This LEMP encodes indeterministic transition rules: the mass might slide down the dome at any arbitrary time in the future, but it is not determined when or whether the slide will occur. But once the mass slides, Newtonian mechanics can be used to predict its trajectory. The physicist would not rob themselves of this predictive power by imposing arbitrary constraints on LEMPs.[19]

[18]Mathematically-minded readers would disagree that Lagrangian and Hamiltonian mechanics assign the same states, for the state-spaces of the two theories have different geometrical structure (one is a tangent bundle, the other a symplectic manifold). For extensive discussion of this issue see: North 2009; Curiel 2014. However, from a prescriptive standpoint, the two theories yield the same equations of motion whenever they are both applicable. The two state assignments are thus *dynamically* equivalent.

[19]The other requirements are similarly discarded in favor of predictive power. The requirement that forces be compliant with Newton's third law (action and interaction) would exclude constraint forces (Wilson

Ruetsche has recently pointed out a similar situation in applications of Quantum Field Theory to curved spacetime. In semi-classical gravity models, where one treats the stress-energy tensor $T_{\mu\nu}$ (but not the Einstein tensor) as a quantum operator, a constraint known as the Hadamard condition must be imposed on the correlation functions of the quantum fields in order to ensure that the expectation value of stress-energy $T_{\mu\nu}$ remains finite (Wald 1994, §4.6). Thus, one might be tempted to say that only Hadamard fields are physical. Nevertheless, as Ruetsche (2015; 2016) has pointed out, the Hadamard condition is not a universal constraint. In dealing with Hawking radiation, for instance, one needs to assign non-Hadamard fields to the system (using the same commutation relations as state assignment rules). Now, the Hadamard condition is a constraint on correlation functions, which ultimately provide the LEMPs. Thus, one can diagnose the issue as follows: the descriptive attributes of semi-classical LEMPs are different from those of Hawking LEMPs. Therefore, any strict requirement that all LEMPs be Hadamard would hamstring QFT in its quest for maximum predictive power.[20]

Of course the idea of local theorizing is not new in itself: Norton, Wilson, and Ruetsche, from whom several of the above examples were borrowed, have championed different versions of the idea for some time (also cf. the local notion of "system laws" in Woodward and Wilson 2019).[21] I am obviously sympathetic to these philosophers. As the examples show,

[manuscript], 5) and arguably electromagnetic forces (assuming a vectorial form of the third law); demanding an identifiable source entity for all forces disqualifies entropy and viscous forces (Wilson 2018, 26-27); the requirement that forces be conservative leaves out dissipative forces; and so on.

[20] ABUNDANCE is in direct contrast to Giere's views (see above), for according to ABUNDANCE *there is a set of necessary and sufficient conditions for what secondary formulae are allowed: namely that they be fitted to local data.* Moreover, ABUNDANCE implies that new models should not be required a priori to have any sort of "family resemblance" to existing ones. In other words, the descriptive attributes of LEMPs are not treated by the scientist as *global* constraints, but as *local* facts of experience. Hence the adjective "local" in "local empirical mediating principles".

[21] Broadly speaking, there have been two species of local philosophy of science: *methodological* and *semantic / ontic*. The methodological variety – i.e. the notion that different areas of science require different heuristics and principles of reasoning – includes Cartwright's "dappled world" (1999), Norton's *material theory of induction* (2016) and Wimsatt's "piecewise approximations" and "heuristic strategies" (2007). The semantic / ontic version of locality – i.e. the idea that theoretical descriptions of the world aquire different meaning or ontological character in different local contexts – includes Norton's arguments against causal fundamentalism (2003; 2008), Wilson's arguments that concepts of classical mechanics often have "wandering significance" which allows them to bridge different "patches" of applied mathematics (2008; 2018), Ruetsche's "locavore" ontology of quantum field theories (2015; 2016), as well as Lloyd's "nesting models" account of evolutionary biology (1994 [1988]; 2013). Of interest to my project is the semantic / ontic variety, which can be easily accommodated in a prescriptive-dynamical framework.

the locality pointed out by them is often (though perhaps not always) rooted in the locality exhibited by LEMPs. This diagnosis strikes me as illuminating, in particular since some of these local accounts have been limited to specific subdisciplines (e.g. Wilson to classical mechanics, Ruetsche to QFT in curved spacetime). Thus, the prescriptive-dynamical view can serve as a broader framework in which to accommodate local philosophies of science.

Nevertheless, I suggest that local philosophies of science have not gone far enough in their rejection of traditonal accounts, for none of the authors have considered the descriptive locality of theories as an invitation to abandon the descriptive-ontological view.[22] The continual adherence to descriptivism is understandable given that no diagnosis has been offered for the origin of descriptive locality and thus the issue has remained somewhat mysterious. I diagnose the root of descriptive locality to be the open-endedness of transition rules which I shall argue calls for abandoning descriptive-ontological construals of theories in favor of a prescriptive-dynamical one.

But first let me point out another closely-related aspect of open-endedness that will prove important for my argument below.

1.4.3.2 Characteristic 2: LEMPs are inexhaustible

ABUNDANCE is a fact about theories, namely that they put no global restrictions on their LEMPs, effectively allowing any predictive LEMP as long as the state assignments are still applicable. But that is also a reflection of a fact about nature, namely that the transition rules tend to surprise us in new circumstances, whether this be a new context of interac-

[22]To be fair, Ruetsche comes closest, for instance in her occasional reference to the cannonical commutation relations as "recipes" (2015, 3437 ff.). Nevertheless, she remains committed to descriptivism, evidenced by her commitment to Fine's Natural Ontological Attitude, which still takes theories as guides to ontology, notwithstanding the emphasis on the *multiplicity* of local ontologies (see ibid, 3426). I should recognize the sole exception at this point: in *The Quantum Revolution in Philosophy* (2017), Richard Healey explicitly argues against descriptive readings of quantum mechanics and uses prescriptive language to interpret quantum mechanical state assignments. Healey's account is the closest to mine, with two differences. Firstly, Healey's are prescriptions of *credences*. To be sure, the credences are derived from the state assignments, which are prescriptions in my view. Nevertheless, Healey is highlighting a second sense in which quantum mechanics is prescriptive, one compatible with but slightly different from mine. Secondly, Healey's prescriptiveness is unique to quantum mechanics, whereas mine is more general. (Prof. Healey has expressed his agreement with the foregoing assessment of the relationship between his view and mine in personal communication.) Nevertheless, Healey provides the essential roadmap for dispelling oneself of (quantum) descriptivism. More on this in §1.6.1.

tion or a new natural system altogether. I shall refer to the fact that new circumstances might require new LEMPs as the *inexhaustibility* of LEMPs because it means that there is no exhaustive list of LEMPs to choose from when one finds oneself in new experimental circumstances.[23]

First consider new contexts of interaction with the same system. An example we have already discussed occurs in modeling a restorative system such as a mass-spring or a pendulum, where $F = -kx$ is a common choice for the second formula in textbooks of Newtonian mechanics as well as in practice. But despite being called "Hooke's law", the formula $F = -kx$ does not have a pre-ordained scope of validity and will cease to apply as soon as the spring begins to behave non-linearly. While any given spring or pendulum has a linear (Hookean) phase, there is a threshold of stretch beyond which non-linear terms in x and eventually $\dot{x}$ begin to show up in the restorative force formula. A new context of interaction (a wider range of stretch, a new temperature, a medium other than air, etc.) might always require a new LEMP to model the spring.

This is not specific to restorative systems of course. Consider how the same liquid has different viscosity functions under different circumstances (temperature, pressure, disturbance, etc.), which supply the local LEMPs of fluid mechanics. Or consider another LEMP of Newtonian mechanics, the inverse-square formula: $F = \frac{Gmm'}{r^2}$. Both Modifiend Newtonian Dynamics (MONDs) and dark matter shift this "law" to fit the data to galaxy-level contexts of experiment. Once again, the scope of validity is not pre-ordained.

In short, the first way in which LEMPs are inexhaustible is that a new context of interaction is always allowed to surprise us.

INEXHAUSTIBILITY I: New contexts of interaction might require new LEMPs.

The second way in which LEMPs are inexhaustible is upon examining *a new system* that has not been studied before. Here too there can be no *a priori* knowledge of what the LEMPs for that system will look like, as the theory does not come with a catalog of physically relevant LEMPs to choose from.[24] Rather, as said above, the theory asks us to

[23] I do not mean to claim that the inventory of the fundamental building blocks of nature itself is inexhaustible (I doubt that we can have such knowledge one way or another). Rather, my claim is that our theories do not tell us when this inventory has been exhausted.

[24] I do not mean to imply that the fixed formulae are "a priori" in Friedman's neo-Kantian sense (2001). The

presuppose the master formula and carefully "reverse engineer" the LEMP. For instance, having presupposed $F = ma$, Newton spent much of the Principia fine-tuning the inverse-square formula to the best of his data. He proved that in general, if $T \propto r^n$, where T is the period of the planet and r the distance to the Sun, then $F = ma$ implies that $F \propto r^{-(2n-1)}$ (Newton 1999 [1687], Prop. IV, Cor. 7). In accordance with this, Newton carefully considered possible deviations from $n = \frac{3}{2}$ in the data, ultimately finding the deviations negligible within the margin of error (ibid, Book III, Theorem 2, Prop. 2; see also Norton 2010 [2002], 128). Following this method, several physicists such as Laplace (1798-1827) and Asaph Hall (1894) attempted to read back corrections of the form $F \propto \frac{1}{r^{2+\delta}}$ from the data, which Hall considered a *continuation*, rather than a *disconfirmation*, of Newton's project. And of course this is not specific to gravity. Much of the project of Newtonian chemistry in the 19th century was devoted to finding the right force formulae to account for chemical reactions. One proposal, contemplated by James Keill among others, was a short-range attractive force that diminishes faster than gravity and depends on the shapes of the particles (cones, cylinders, cubes, spheres) as well as their masses (see Thackray 1970, 67 ff.). The attitude exhibited in these research programs confirms the scientist's tacit commitment to the inexhaustibility of LEMPs for new systems (and as a result, lack of commitment to any sort of "family resemblance" a la Giere).[25]

INEXHAUSTIBILITY II: New systems might require new LEMPs.[26]

---

master formulae are not conditions for the possibility of empirical knowledge, although they are instructions for constructing LEMPs.

[25]The same goes for non-classical theories. In Quantum Field Theory, for instance, one holds the canonical commutation relations of fermionic and bosonic fields fixed, and experimentally reads off interaction Lagrangian terms of the form $\mathcal{L}_{\text{int}} = -ie\bar{\psi}\gamma_\mu A^\mu \psi$ for various species of particles and the interactions they participate in. Thus, there is no theoretical prediction for the number, form, or strength of terms in the interaction Lagrangian: the number of generations of fermions and the existence of right-handed neutrinos, for instance, are open-ended in the Standard Model. Each interaction term corresponds to a particular set of experimental findings. This means that if new transition amplitudes – e.g. corresponding to some of the many versions of the inflaton field – turn up in future experiments, we must simply accept them into the family. However, the Standard Model does present a potential trouble case for my view, insofar as the Weinberg-Salam electroweak Lagrangian (a LEMP) *predicted* the existence of the Higgs field. A satisfactory treatment of this case is beyond this paper, but my short response is as follows: sometimes the state assignments plus some empirical data imply certain *coherence conditions* which essentially state that either the transition rules must obey certain constraints or else the state assignments lead to contradiction. It is therefore not precisely *theoretical* considerations that fixed the electroweak Lagrangian, but *logical and empirical* ones, namely the fact that the previous LEMPs required gauge invariance for the theory to be coherent, and guage invariance required the Higgs field.

[26]What if the "new system" can be modeled as composed of several component systems with known

One might object to INEXHAUSTIBILITY by saying that we *do* know when Hooke's law applies, namely it only applies to "small" disturbances in restorative systems. After all, once we interpret $F = -kx$ as the first term in a Taylor expansion $F = \Sigma_n \frac{f^{(n)}(0)}{n!} x^n$, we can delineate the scope of Hooke's law as the range of displacements in which the higher order terms are negligible. My response is that this does not help us delineate the scope of validity of Hooke's law. First of all, note that *any* arbitrary restorative system will conform to Hooke's law for *some* range of displacements. Therefore, "small" disturbances in this context are simply defined as "those displacements for which higher-order terms of the above expansion are negligible", which is equivalent to "those displacements that conform to Hooke's law". In other words, this is equivalent to saying that Hooke's law is valid as long as Hooke's law is valid! Thus, the scope of validity of Hooke's law is *circularly defined*. And similarly for LEMPs in general.

Corollary to INEXHAUSTIBILITY: The scope of validity of a LEMP is circularly defined.

The above also helps respond to another objection to the idea of inexhaustibility.[27] According to this objection, the appearance of open-endedness stems from how one individuates "theories". If one identifies one big "Newtonian theory" and considers different Newtonian systems as models of this umbrella theory, then it looks like the theory is open-ended. But alternatively, one can identify $F = ma$ as a theory *fragment* and its conjunction with each LEMP as a separate, complete theory with a specific domain of application: Newtonian gravitational theory, Newtonian projectile theory, Newtonian pendulum theory, Newtonian strut theory, and so on. Then, the objection goes, each of the mini-theories can be considered fixed *within its own domain*.

There are several issues with this approach. First of all, it is not clear that this approach amounts to more than a verbal maneuver. After all, the mini-theories are not handed down

LEMPs? Could one deduce the LEMP of the total system from those of the components? I do not deny the possibility of educated guesses about the LEMP of the total system; but these guesses inevitably hinge on certain substantive assumptions about compositionality / additivity of the components' LEMPs. However, compositionality and additivity are ultimately empirical assumptions, i.e. it is ultimately up to experience to decide whether LEMPs are holistic or not. As such, it would not be a violation of Newtonian mechanics if the force formula for a three-body gravitational system were different from the sum of pair-wise gravitational forces.

[27] Thanks to an anonymous referee for discussion of this objection.

from the sky: when an engineer creates a new elastic band or metal beam with novel properties, they still need to construct a new "mini-theory" for this new system, and they must do so under the guidance of the master formula $F = ma$ combined with local experimental data. That is of course why the master formula features in each of the mini-theories as an "axiom". As such, the list of mini-theories is still open-ended and context-dependent; there is still a guiding principle, and this principle is still generating an open-ended list of predictive equations through engagement with the system. It therefore appears that we have achieved a resemblance of fixity merely by insisting that one refrain from using the word "theory" in reference to the open-ended guiding principle, which is merely a verbal difference.

Secondly, it is not true that the mini-theories are fixed within their own domains. For instance, if "Newtonian gravitational theory" specifically refers to the conjunction of $F = ma$ and $F = \frac{Gmm'}{r^2}$, then Newtonian gravitational theory is not fixed within its own domain. If it were, Laplace and Hall would have considered their work on possible modifications to the inverse-square formula a potential *refutation* of Newton. But as explained above, they considered theirs a continuation of Newton's project, which implies that they considered the inverse-square formula open-ended *within the domain of celestial mechanics*.

Thirdly, as the Corollary above implies, since each "domain" is defined by a fixed LEMP, the boundaries that separate the mini-theories are circularly defined. Again taking "Newtonian gravitational theory" to refer to the conjunction of $F = ma$ and $F = \frac{Gmm'}{r^2}$, one can see that there is no sharp boundary for when this mini-theory is valid. For example, is motion at the scale of galaxies included in the domain of application of Newtonian gravitational theory? It is according to Dark Matter theories, which keep the same LEMP $F = \frac{Gmm'}{r^2}$ and merely posit invisible mass in the system. By contrast, according to Modified Newtonian Dynamics (MOND), galaxy-scale motion *is* a new domain for Newtonian gravitational theory. This is because a "domain" is defined as those applications for which the LEMP is fixed, and MOND recommends a modification of the LEMP for galaxy curves. But this is once again circular: the validity of the LEMPs determines the domain in retrospect, while the domains are themselves supposed to tell us when each LEMP is expected to apply and when it needs to be changed.

This concludes our preliminaries on LEMPs. I will now briefly sketch the prescriptivist

alternative that is suggested by the considerations above before moving on to argue against descriptivism.

1.5 The Prescriptive-dynamical Construal in a Nutshell

The above considerations suggest the following prescriptive-dynamical account of the essential content of physical theories: a physical theory is a prescriptions for state assignment to natural systems for a range of interactive contexts. More precisely, the prescriptions of the theory come in three steps:

1- (Fixed) Assign dynamical states of such and such a form to the system (e.g. $(x, \dot{x})$).
2- (Open-ended) Read off the transition operator $D_{\delta t}$ from local experimental phenomena.[28]
3- (If possible) Compactify $D_{\delta t}$ in a simple quantity (e.g. F, L, or H).

The two main tenets of the prescriptive-dynamical view therefore are: i) the essential content of the theory must be couched in dynamical terms (states and transition rules) rather than in ontological terms (entities, properties, structures), and ii) the fixed part of the theory's working posits is an injunction to interface with the system through certain state assignments, rather than a description of what is there and what it does.

The prescriptivist thus rejects the widespread attitude in philosophy which Woodward, following Dewey, has labeled "the spectator view" (Woodward 2020 [preprint]): an approach to epistemology in which *passive observation* is taken as the paradigm example of knowledge acquisition. It is natural for a philosopher taking this approach to consider a theory a description of what one is looking at. If, by contrast, one takes *active interface* as the paradigm of knowledge acquisition, one is naturally led to the idea that a theory tells us *what to do*, rather than *what there is*. As such, the prescriptive-dynamical view has strong affinities with pragmatism, though I shall not insist on this label and its myriad connotations in what follows.

Let us now turn to my argument against descriptivism.

[28]Note: I use "transition" in a more general sense than change in time. The "transitions" of Poisson's equation ($\nabla^2\phi = f(x, y, z)$), for example, are purely spatial. Thus, the transition operator may well be $D_{\delta x}$, etc.

1.6 The Descriptivist's Dilemma

In this section, I will summarize how the foregoing considerations rule out all forms of descriptivism. I shall begin by examining the possibility of descriptive content in the *fixed parts* of theories, which I argue is extremely thin if at all existent. I will then proceed to show that attempting to pinpoint the descriptive content of the *open-ended parts* traps the descriptivist in an unpleasant dilemma.

1.6.1 Do the fixed formulae have descriptive content?

My concern in this paper is with the *essential content* of physical theories, a notion akin to Kitcher's *working posits* (1993, 149 ff.) and Psillos's "indispensable" or "essential contributors" (1999, 108 ff.). The essential content would be those parts of a theory that actually contribute to its predictive power as opposed to any idle baggage that might come with the theory. My claim is not that physical theories in their many formulations throughout history have never contained any descriptions; to the contrary, physicists often use descriptive statements to scaffold their theories and make them more intuitive and visualizable. My claim, rather, is that insofar as physical theories contain descriptive statements, these play no role in the predictive power of the theory.

A teacher of Newtonian mechanics might start by describing a world full of point particles and rigid bodies and forces that push them around, proceed to claim that these objects carry properties such as mass, momentum, etc. before finally writing down the equation: $F = ma$. However, once said equation is written down, the only relevant manipulations the student needs to retain are those summarized in §1.5 above, namely: how to assign a state consisting of a position x and a velocity $\dot{x}$ to the system of interest at any given time, and how to pack/unpack the transition rules in the quantity F by forming and solving second-order differential equations. The *essential* insight of classical mechanics is simply that the state of the system must be tracked through the ordered pair of position and velocity, not position alone, and not any other combination of measurable quantities. That insight is a prescription, not a description.

As we have seen, Newton's theory does not tell us whether forces are deterministic or not, whether they must emanate from an identifiable source object or not, whether they conserve energy or not, etc. Nor does the theory tell us what variables the force can be expected to depend on (position, velocity, mass, shape, size, composition, color of the object, the age of the particle, ... ?). The problem persists even for a structuralist, for the mathematical structure of the state space changes with the choice of LEMP (see, e.g.: Swanson and Halvorson 2012, 9). For instance, some but not all Lagrangians and Hamiltonians impose a metric structure on the state spaces of their respective theories (see: Barrett 2015). In general, the descriptively salient content of the theory largely depends on the open-ended LEMPs and appears quite empty without them. Without the local force formulae, for instance, the descriptive content of $F = ma$ is simply "the product of mass and acceleration will equal *something*", which is trivially true of any system that can be assigned mass and acceleration, and false otherwise.[29] So $F = ma$ picks out a class of natural systems (namely those that can be assigned mass and acceleration), but the theory makes no essential statements about which systems allow these assignments and which do not; nor does it provide any description of the further characteristics of those systems that do.

But one might still insist: doesn't Newton's theory at least make such claims as that there are forces in the world that move objects around, even if it does not tell us much about the attributes of these forces? My answer is once again: such claims are not essential to the predictive power of the theory. As explained above, the essential core of Newtonian mechanics is a differential equation of motion of the form $\ddot{x} = f(x, \dot{x}, t, ...)$. It is this equation that takes us from initial to final states, thus enabling us to make predictions. But forces do not feature in the equation of motion, because they are eliminated from the two formulae that yield said equation. Thus, forces are certainly not implied by the immediate success-generators of Newton's theory. Consequently, the only way the existence of forces would be considered essential to Newton's theory is if the *derivation* of equations of motion *had to* go through the assumption of forces. But this is clearly not the case, as one can often derive the same equation of motion from a force-free theory such as Lagrangian mechanics.

[29]Indeed, the sentence in quotes is the "Ramsified" version of Newton's second law, which is precisely what structuralists such as Sneed (1979 [1971]) and Stegmüller (1979) considered to be the entire content of the theory. Ironically, they, too, puzzled over the triviality of this statement.

Since forces are neither featured in the predictive equation of motion nor necessary for the derivation of the same, they are not part of the essential content of the theory.[30]

Note that I am not making an underdetermination argument. My claim is *not* that we should doubt forces because they have alternatives, but rather that forces fall outside the essential content of the theory, and are thus immune from underdetermination problems. It is precisely the supposition that forces are part of the *essential* content of Newtonian mechanics that leads to underdetermination. For if forces are essential, then presumably so are Lagrangians, Hamiltonians, etc. And once both forces and Lagrangians, say, are considered essential, one has set oneself up for a powerful underdetermination argument: two theories with radically different essential contents are seen as saving the same empirical phenomena, thus raising questions about believing their essential posits.

The prescriptive-dynamical view, on the other hand, avoids this underdetermination problem entirely, for the latter takes forces and Lagrangians to be descriptive scaffolding: alternative ways of compactifying what is essentially the same local transition rules. As we saw in §1.4.3.1, compactification rules are used opportunistically and chosen based on pragmatic considerations. The common content of all these formalisms is the prescriptions for forming the differential equation of motion, which only specify state-assignment and transition rules: forces and Lagrangians are ultimately eliminated in order to derive this equation. If anything, this argument should help us dodge the underdetermination problem by separating the *essential, dynamical* content of the theory from its *inessential, descriptive* baggage. Rather than arguing that we should not believe in forces despite the enormous predictive power of the theory that posits them, I argue that the predictively relevant part of the theory posits no such things.[31]

[30]This argument is independent of where one falls on the debate about the (in)equivalence of Newtonian, Lagrangian, and Hamiltonian mechanics, for the point can be made about a specific system or group of systems to which more than one formulation of classical mechanics applies.

[31]How about the existence of masses or various families of particles? For instance, doesn't the Standard Model of particle physics contain the claim that there are massive vector bosons such as the Z and W? To see if this is the case, first of all note that the Standard Model is a LEMP in my terminology for which the canonical commutation relations of quantum field theory operate as the fixed master formulae. Quantum field theory in itself does not imply the existence of anything; it merely lays out a number of ways one could assign quantum field operators to systems. Thus, even if the Standard Model implies the existence of various kinds of fields / particles, that is consistent with the prescriptive-dynamical framework, insofar as the latter locates descriptive content in the local LEMPs.

But regardless, the appearance that each kind of quantum field represents a specific kind of entity is

I conclude that the fixed formulae contain extremely thin descriptive content, if any at all.[32]

1.6.2 The descriptive content of open-ended formulae: a dilemma for the descriptivist

As we have seen, the LEMPs are needed to give theories non-trivial descriptive content. But the LEMPs are numerous and each might carry some novel descriptive element. The descriptivist might hope to reconcile these two facts by simply conceiving of the theory as either a massive disjunction or a massive conjunction, with each disjunct / conjunct corresponding to one choice of LEMP. The two cases are slightly different, but ultimately neither works, because the LEMPs are inexhaustible, and thus one can never know if the disjunction or the conjunction is complete. More precisely, INEXHAUSTIBILITY imposes an unpleasant dilemma on the descriptivist: should the disjunction or conjunction include all *possible* LEMPs, or *known* LEMPs only? As I will argue, neither horn is pleasant for the descriptivist.[33]

Consider the disjunctive case first, which is simpler. The suggestion is to formulate the

misleading anyway. A massive vector boson, for example, is replaced through the Higgs mechanism with two *massless* bosons (one vectorial and one scalar), without affecting the resulting equation of motion. If descriptive claims that such and such particles of such and such masses exist were essential to the Standard Model or quantum field theory, either the same equation of motion could not be derived without them, or else one would face a vexing underdetermination problem.

Once again, this is unproblematic from a dynamical point of view, for the latter construes each type of field operator (scalar, vector, spinor, etc.) as instructions for dynamical state assignments, which instructions can be scaffolded with a number of different ontologies. It is not surprising that a dynamical state with three degrees of freedom (massive vector) can be swapped with one state with two degrees of freedom (massless vector) plus one with one degree of freedom (real massless scalar). The ontologies differ radically, but the state assignments are equivalent, and underdetermination is avoided. (A detailed treatment of how quantum operators can be treated as prescriptions for state assignment is explored in Healey 2017.)

[32]To be clear, the fact that a certain state assignment is the best prescription for a given context *implies* a descriptive fact about how the world works, namely something to the effect of "the (hidden) underlying dynamics of the system is such that these state assignments *track* the true transitions of the system." But this is hardly much of a description at all, and certainly not comparable to ordinary descriptive statements such as "there are particles and forces and they are related in such and such a way," which the prescriptivist considers as either embellishment or scaffolding for the state space.

[33]One can extend the argument to an account with movable pieces (such as Giere's) by asking: should one include all possible family resemblances or only known ones? Since family resemblance is not a local or empirical matter, Giere's view faces a similar dilemma as the traditional, fully-fixed accounts.

theory as a massive disjunction:

$$(L\&M_1) \vee (L\&M_2) \vee ... \equiv L\&(M_1 \vee M_2 \vee ...) \tag{1}$$

where the propositional constant L denotes the "law-like" master formula and M_i the various mediating principles (LEMPs). But this runs into a dilemma: should one include all *possible* LEMPs among the M_i or known LEMPs only? Neither option works.

Including all possible LEMPs makes $(M_1 \vee M_2 \vee ...)$ a tautology and reduces the theory down to L. This cannot be right. L is not a predictive equation and thus cannot be the entire content of the theory. At least not if "the theory" is to be something that figures in a predictive schema such as Hempel's (see Section 1.3), which may be summarized as follows:

The theory
Auxiliary assumptions
Correspondence rules
Initial / boundary conditions
———————————
Prediction

As the above schema makes clear, "the theory" cannot simply be the master formula alone or else the predictive schema would be deductively invalid. As argued above, the master formula is fairly vacuous on its own.

On the other hand, including known LEMPs only would mean that all the disjuncts are false for any *new* context which might require a new LEMP, even if the master formula is still applicable in that context. But according to INEXHAUSTIBILITY, such contexts are always possible, and according to ABUNDANCE, *some* LEMP will predict the behavior of the system in such contexts. Consequently, including known LEMPs only would merely hamstring the theory in making predictions about new circumstances. Therefore, the disjunctive account fails under both horns.

Let us consider the conjunctive view next. Presumably, the suggestion is to consider the theory as a conjunction of statements of the form: "If such and such circumstances obtain, use such and such a LEMP". The full theory would thus look like the following:

$$[C_1 \rightarrow (L\&M_1)]\&[C_2 \rightarrow (L\&M_2)]\&... \tag{2}$$

where C_i are the descriptions of the circumstances under which each LEMP is valid. There are two problems with this. First of all, the circumstances C_i are ill-defined. This is due to the corollary to INEXHAUSTIBILITY. The circumstances C_i are meant to delineate the scope of validity of each LEMP M_i, but according to the corollary, such scopes can only be defined circularly, i.e. we have $M_i \rightarrow C_i$. Given this, the ith conjunct is logically equivalent to $C_i \rightarrow (L\&C_i)$, which is the same as $C_i \rightarrow L$. Thus, the theory as a whole is given by $(C_1 \rightarrow L)\&(C_2 \rightarrow L)\&...$, which is in turn logically equivalent to $(C_1 \vee C_2 \vee ...) \rightarrow L$. But this cannot be the content of the theory, for this proposition merely states the scope of validity of the master formula L: it says that as long as one of several conditions holds, the master formula applies. This is not an empirically predictive proposition, because as said above the master formula by itself does not make any predictions.[34]

But let us put aside the circularity of the conditions. Consider the original massive conjunction again and assume the C_i are well-defined. What C_i should we include? All possible circumstances or only known circumstances? Once again, neither option works. For to include all possible C_i, one would need to know the corresponding LEMP for every possible circumstance, which we do not know according to INEXHAUSTIBILITY. Moreover, unless one is dealing with a "theory of everything", the master formula is bound to become inapplicable under some possible circumstances, which means that we cannot include all possible C_i. Thus it might seem reasonable to include only those C_i under which the theory is known to work. But this means that given any new circumstance C' that was not among the original C_i, all conjuncts of the theory will be vacuously true, and the theory will be contentless with no predictions. Once again this is problematic because of INEXHAUSTIBILITY and ABUNDANCE, as this approach would deprive the theory of its ability to make predictions in novel circumstances. Therefore, the conjunctive account also fails.

[34]Besides, what circumstances are the C_is meant to range over? If they are meant to cover *all possible circumstances*, then $(C_1 \vee C_2 \vee ...)$ is vacuously true, and the massive conjunction simply boils down to L itself, which once again is not the entire content of the theory. If, on the other hand, the C_i are only meant to range over known physical circumstances in which the theory has been tested, then all of the conditionals become vacuously true in all *untested* circumstances, which means that the theory is empirically impotent outside of known circumstances. But again, INEXHAUSTIBILITY means that new LEMPs are always possible, and ABUNDANCE implies that the theory *can* make predictions about new systems and contexts as long as its master formula is still applicable. Including only known LEMPs would once again deprive the theory of future success in any amenable contexts that have not been encountered before.

A final note: the above argument also indicates that prescriptivism is the only way to prevent the master formula from becoming dispensable. Both disjunctive and conjunctive accounts make the master formula superfluous, because each pair of statements $L\&M_i$ owes its predictive power to a differential equation from which the theoretical terms of L and M_i have been eliminated. For example, if the theory is Newtonian mechanics (taking the disjunctive account for simplicity), we are told to write the theory as

$$[(F = ma)\&(F = -kx)] \vee [(F = ma)\&(F = cx^{\frac{1}{2}})] \vee .... \tag{3}$$

But then one could eliminate the force from each pair of equations and simply write:

$$(\ddot{x} = -\frac{k}{m}x) \vee (\ddot{x} = \frac{c}{m}x^{\frac{1}{2}}) \vee .... \tag{4}$$

Assuming the disjuncts are fixed (even if infintie), there would be no need to repeat a formula such as $F = ma$, which contains an eliminable unkown. Why not simply a disjunction of differential equations which are couched in purely experimental terms? Why does the scientist insist on keeping the master formula around? The descriptivist struggles to answer this question[35]. According to prescriptivism, on the other hand, we keep the master formula around because it provides a *recipe* or *prescription* for interacting with nature in new contexts and systems.

[35]This is a manifestation of Hempel's Theoretican's Dilemma. Thus, the prescriptive-dynamical view provides a new solution to the Theoretician's Dilemma, one that unlike Psillos' (1999, 22 ff.) does not require questioning the observational-theoretical distinction itself. But that is beyond this paper.

1.7 Conclusion

I have argued that the essential content of the most foundational theories of physics consists of fixed prescriptions for what quantities to keep track of (the dynamical states of the system), and what quantities to record (the transition rules). These prescriptions are by themselves devoid of descriptive content, but they are "filled" with open-ended empirical formulae that encode the local transition rules. The latter are unrestricted by theory (ABUNDANCE) because they are local and empirical, and tend to take new and surprising forms in new contexts (INEXHAUSTIBILITY). This open-endedness means that the descriptive content of the theory cannot be formulated as a massive disjunction or conjunction, as the descrip-tivist would have it. Insofar as theories describe entities, properties, and structures, such descriptions are *inessential posits* (scaffolding, embellishment, etc.) that do not contribute to the success of the theory.

2.0 Chapter 2: Theoretical Equivalence as Dynamical Equivalence

2.1 Introduction

In this chapter, I will discuss the implications of my view for theoretical equivalence. The main operative premise for the majority of my arguments below – which I take as having been established in Chapter 1 – is the following dual fact: *Predictive physical theories have fixed state assignments and open-ended transition rules.* I called this view the *prescriptive-dynamical view*, because the fixed parts are prescriptions for constructing the open-ended parts from local data. I will begin in Section 2.2 by examining the notions of theoretical equivalence that emerge from this prescriptive-dynamical construal of theories. The main one that I argue is relevant to the practice of physics is what I call "dynamical equivalence". Section 2.3 then surveys three mainstream accounts of theoretical equivalence from the current literature and raises some general objections to these definitions. I will then proceed in Section 2.4 to a critical review of the recent debates on classical mechanics, which have arisen from applications of mainstream accounts of theoretical equivalence to the case of classical mechanics. I shall argue that these debates are founded on a misconstrual of physical theories and their two-tier, fixed/open-ended character, and that the bulk of the disagreement and confusion in this literature can be resolved by recognizing that dynamical theories have fixed state assignments and open-ended transition rules. This discussion will also illustrate the application of prescriptive-dynamical notions of equivalence to the case of classical mechanics. Finally, Section 2.5 presents evidence that my interpretation of theoretical equivalence as dynamical equivalence explicitly conforms to that of the practicing physicist by reviewing some quotes from textbooks of classical mechanics.

2.2 Theoretical Equivalence from a Prescriptive-Dynamical Point of View

2.2.1 Theoretical equivalence and essential content

Two theories are said to be *empirically equivalent* when they make the same predictions about observable or measurable facts. This is to be distinguished from *theoretical equivalence* (the topic of this chapter) which is a stronger condition and occurs when two theories' unobservable parts agree.

Now, one would not want to demand that *all* unobservable aspects of the two theories match for them to count as equivalent. What matters is rather *the essential content* of the two theories. Following others in the literature, I use "essential content" to refer to that (hypothesized) core set of assumptions of a theory that plays a role in its empirical success. The essential content is to be contrasted with the "inessential" or "idle" parts of the theory, if any, which are typically mere embellishments and/or scaffolding and do not contribute to predictive success. (See Chapters 1 and 3 for a more detailed discussion of essential content.) A reasonable standard of theoretical equivalence would allow equivalent theories to disagree on the inessential components, but require that they converge on their essential posits.

Since theoretical equivalence means sameness of essential content, any answer to the question "What is the essential content of a theory?" will inevitably serve as the basis for an answer to the question "When are two theories equivalent?" and vice versa.[1] In Chapter 1, we answered the first question roughly as follows: the essential content of a physical theory is a local prescription for assigning dynamical states to the system and recording the transitions of these states. Let us briefly recap this account and see what notion of equivalence we can infer from it.

2.2.2 Some basic facts from Chapter 1

As explained in Chapter 1, the prescriptions of a theory break down into three commands:

1- State-assignment rules (fixed): Assign dynamical states of such and such a form to the system (e.g. $(x, \dot{x})$).

[1] Barrett 2017, 22-23 makes extensive arguments for this proposition. And see Coffey 2014 for a similar conclusion. Also cf. Nguyen 2017, 982.

2- Transition rules (open-ended): Read off the transition operator $D_{\delta t}$ from local experimental phenomena.[2]
3- Compactification rules (on an if-possible basis): Encode $D_{\delta t}$ in a compact quantity (e.g. F, L, or H).

The first step is the state assignment rules, which are shown schematically here. This part of the theory tells us what collection of measurable variables should be tracked as the dynamical state of the system, and how these assignments should be coordinated in space and time. In classical mechanics, these variables are the position x (or q for generalized coordinates) and the velocity $\dot{x}$ (or $\dot{q}$ for generalized velocity), which must be coordinated using the transformations of the inhomogeneous Galileo group.

The second step is the empirical measurement of local transition rules. In this procedure, the scientist must *presuppose* step 1 and "fill" the prescription by engaging the system in accordance with the state assignment rules and consequently recording the evolution operator that arises from the local empirical data.

Thirdly and finally, the scientist is asked to encode the obtained transition rules into a compact quantity. In classical mechanics this may be a force function, a Lagrangian, or a Hamiltonian. This will give us the *local empirical mediating principles (LEMPs).* This should all be familiar from Chapter 1.

2.2.3 Prescriptive-dynamical notions of theoretical equivalence

What account of theoretical equivalence can we derive from the above? Since there are three nested steps to the prescriptions of any given theory (each step presupposing the last), on a crude examination one would expect three ways in which theories can be equivalent or inequivalent, corresponding to agreement or disagreement on each of the three steps:

Equivalence I: The two theories agree on step 1 but disagree on step 2 (and therefore 3).
Equivalence II: The two theories agree on steps 1 and 2 but disagree on step 3.
Equivalence III: The two theories agree on all three steps.

[2]Note: I use "transition" in a more general sense than change in time. The "transitions" of Poisson's equation ($\nabla\phi = f(x, y, z)$), for example, are purely spatial. Thus, the transition operator may well be $D_{\delta x}$, etc.

However, in reality Equivalence I is not a real possibility. Step 2 is local and open-ended: it is "reverse engineered" from experiment by presupposing the state assignments. Therefore, the (in)equivalence of two theories cannot possibly hinge on (dis)agreement on step 2. Rather, either of two things can happen: i) two theories agree on state assignment rules, in which case they are bound to agree on transition rules, for these are inferred from local data; or ii) they do not agree on state assignment rules, in which case they cannot agree on transition rules, for transition rules are relations between states and defined in terms of them.

Therefore, there are at most only two types of prescriptive-dynamical (in)equivalence: one hinging on (dis)agreement on step 1 and the other on (dis)agreement on step 3. Let us call these two notions Dynamical Equivalence and Compact Equivalence, respecitvely.

2.2.3.1 Dynamical equivalence

Let us call two theories *dynamically equivalent* if they agree on step 1 (and therefore 2).

> *Dynamical Equivalence:* Two theories are dynamically equivalent iff they prescribe the same experimental state assignment rules.

Note that to be equivalent, the two theories must agree on state assignment rules *in experimental practice*. The qualifier "in experimental practice" is very important, because "agreeing on state assignment rules" does not mean having isomorphic geometrical state-spaces or anything on that level of abstraction. As we shall see in Section 2.4, much has been made of the fact that the geometrical structure of Lagrangian and Hamiltonian state-spaces are distinct (non-isomorphic) and that therefore, the argument goes, they are not equivalent in any interesting sense. However, a good pragmatist would only be interested in those aspects of state assignments that prove important for step 2, namely the physical movements that the scientist must go through in the lab in order to prepare or detect (i.e. assign) the natural system in one of the states prescribed by the theory. Thus, the idea is that two dynamically equivalent theories "induce the same movements" in the lab. Dynamical equivalence is the pragmatist's notion of equivalence.

I will further argue that *dynamical equivalence is indeed what the practicing physicist*

means by "equivalence". But first let us examine the other notion of equivalence that arises from agreement on step 3, i.e. the compactification rules.

2.2.3.2 Compact equivalence

Let us call two theories *compactly equivalent* if they are dynamically equivalent and have fully compatible compactification rules such that every transition rule that can be encoded in one can also be encoded in the other.

> *Compact Equivalence:* Two theories are compactly equivalent iff they are dynamically equivalent and have interchangeable compactification rules.

The Schrödinger and Heisenberg pictures of quantum mechanics may be said to be compactly equivalent: they generate the same equations of motion and any transition rule that can be encoded in one can also be encoded in the other, as both theories compactify their LEMPs in a Hamiltonian operator. Hamiltonian and Hamilton-Jacobi theories may serve as another pair of compactly equivalent theories. Of course I do not deny that compactly equivalent theories may nevertheless have different computational virtues.

On the other hand, the same set of state assignment rules (step 1) might be compatible with several different compactification instructions (step 3), and *some compactification rules might simply fail to apply to certain specific systems*. The LEMPs of certain non-conservative systems, for instance, cannot be encoded in a Hamiltonian function (Curiel 2014, 309). As is well-known, there are also non-conservative and/or nonholonomic Newtonian systems that have no Lagrangian or Hamiltonian model.[3]

Unlike dynamical equivalence, which I argue is the practicing physicist's notion of theoretical equivalence, *compact equivalence is of little relevance in the practice of physics*. Let us see why this is.

[3]In general, there is no unique set of necessary and sufficient conditions for Newtonian systems to have Lagrangian counterparts, though one can find multiple sets of sufficient conditions. See Butterfield 2006.

2.2.3.3 Dynamical equivalence is the only pragmatically relevant notion of theoretical equivalence

Compact equivalence does not respect the open-endedness of transition rules: the compactification rules in step 3 often put additional (non-local) restrictions on the transition rules, and are therefore at risk of failing in certain local contexts. As we shall see below in the case of classical mechanics, for instance, Lagrangian and Hamiltonian frameworks fail to capture certain types of acceleration patterns, in which case the physicist typically defaults back to a very liberal notion of "force" that is essentially read back from the product of mass and acceleration. The point is that practicing physicists, who are often sharply aware of the open-endedness of transition rules, do not take step 3 as an essential part of a theory, but rather as an opportunistic set of instructions for simplifying the equation of motion – hence the "if-possible" clause in step 3. (I will present textual evidence for this claim in Section 2.5 below.)

On the other hand, well-formed and manipulable mathematical structures that theoreticians and philosophers typically refer to as "theories" generally come with *both* state assignments and transition operators. From that perspective, a "theory" without transition rules appears as a mere theory-*fragment*. Thus, mathematicians and philosophers who do not dabble in open-endedness might be inclined to argue that compact equivalence is the only relevant notion of theoretical equivalence.

This disagreement often manifests itself in the form of a terminological mismatch in which practitioners of physics tend to use "theory" to refer specifically to the state assignment rules alone, while the mathematically-minded often refer to a fixed conjunction of *both* state assignment rules *and* compactification rules as "theory". The former group would say two theories that agree on step 1 but disagree on step 3 are equivalent (or even further: that we are dealing with only one theory altogether), while the latter group would disagree. As we shall see below, this is indeed the case when it comes to classical mechanics, and it does indeed explain part of the confusion in the literature. The former use of the term "theory" seems to underlie physicists' ubiquitous references to the three versions of classical mechanics as "different formulations of *the same theory*" – an assertion that baffles philosophers

and mathematicians whose stricter notion of "theory" identifies different version of classical mechanics as truly distinct and even incompatible theories.

To avoid arguing over terminology, in the following I will allow both uses of the term "theory", and will simply remain vigilant of the two uses and flag the relevant meaning when important. When "theory" is used to mean the state assignment rules and nothing else, there can be only one of kind of theoretical equivalence, namely dynamical equivalence. On the other hand, when "theory" is used to mean a set of state assignment rules plus a set of compactification rules, another notion of equivalence presents itself, namely compact equivalence.

Terminological confusion aside, the important point to note is the following: compact equivalence is of little relevance to the practicing physicist because it goes with a restrictive notion of "theory" that violates the open-endedness of LEMPs. In practice, there need not be any one-to-one correspondence between the models of two dynamically equivalent theories (taking each model to correspond to one choice of LEMP). The open-endedness of LEMPs means that it is always possible to come across a new system or context of interaction whose transition rules can be compactified in one way but not another. As such, for the practicing physicist, one-to-one correspondence between models (as compact equivalence entails) can never be a strict requirement for declaring two theories equivalent!

We shall see how this fact dissolves much of the recent debate on theoretical equivalence in classical mechanics. But before summarizing this debate, I would like to address one lingering issue: the definition of dynamical equivalence provided above may be criticized as rather vague: what does it mean to "prescribe the same experimental state assignment rules"? How can one prove this relationship mathematically? To remedy this, I shall provide a more rigorous but equivalent definition of dynamical equivalence in the following section, and for that we need to turn our attention to equations of motion.

2.2.3.4 Redefining dynamical equivalence through equations of motion

As discussed in Chapter 1, steps 1 and 2 of a dynamical theory can be captured in a single differential equation. In classical mechanics, this equation is of the form $\ddot{q} = f(q, \dot{q}, t, ...)$,

where the lacuna indicate that f might depend on any number of different quantities. The left-hand side of this equation tells us to keep track of q and $\dot{q}$ as our initial conditions, no more and no less, while the right-hand side tasks us to search for a function f that captures the transition rules. Those are precisely steps 1 and 2. What this equation does *not* tell us is the third step, namely how to compactify the transition rules. Newtonian mechanics asks us to write the equation in terms of Cartesian coordinates and encode the resulting f directly in a vectorial "force" function $F = mf$. Lagrangian and Hamiltonian theories allow for cleverer compactifications that may not be feasible in all cases but that greatly simplify problem solving in the right sort of context (e.g. Lagrangian when there are unknown constraint forces and Hamiltonian when there are cyclic coordinates or when numerical methods are required). The same is of course true of all other physical theories: the first two steps are captured in a differential equation that tells us what to keep track of and what to look for, whereas the third step is a separate command regarding how to compactify the results.

Particular instances of $\ddot{q} = f(q, \dot{q}, t, ...)$ (i.e. with f filled in) are called *equations of motion* (EOMs). One familiar example would be $\ddot{x} = -\frac{k}{m}x$. These equations are obtained by conjoining the master formula (e.g. $F = ma$) with a LEMP (e.g. $F = -kx$) which encodes the missing ingredient for transition rules.[4] The ultimate purpose of both the master formula and the LEMP is to merge into an EOM, which is the sole source of the theory's predictive power. Thus, one might think of physical theories as compactified equation-of-motion-generating schemes (provided the EOMs are understood in a prescriptive-dynamical fashion).

Equations of motion are the closest theoretical items to direct empirical evidence, given that theoretical terms such as force or Lagrangian do not appear in them. Nevertheless, EOMs contain more than "purely observational statements", even if one were to use a liberal notion of "observation" that includes all things measurable. This is because the same set of observations / measurements can be captured in different differential equations. That is, different equations of motion can predict the same trajectory. For instance, the forward

[4]Note that "equations of motion" is sometimes used to refer to the master formula itself, such as Newton's second law or the Euler-Lagrange equations. I shall always use the phrase to refer to the equation that results from putting the master formula and the LEMP together.

motion of a mass-spring system can be captured in the differential equation $\ddot{x} = -\frac{k}{m}x$, but it can also be equally well captured as $\dot{x} = \sqrt{\frac{k}{m}(a^2 - x^2)}$, where a is the maximum stretch. The two equations agree on all observations of a forward-going mass-spring, but they disagree on which aspects of this motion should be encoded as the state of the system and which aspects as the transition rules. Such alternatives instruct us to keep track of variables other than the $(q, \dot{q})$ pair, and thus result in different LEMPs. It is therefore possible to agree on empirical results despite generating different EOMs. This should not be surprising, since the prescription that *such and such collection of variables should be tracked*, which is what the EOM dictates, is not an observable / measurable statement; it is rather a conjecture that forms the central injunction of the theory. The equations of motion are therefore "one level above" pure observation (arguably, the only "pure observation" would be a recording of all the variables of interest at all times and places).

As such, equations of motion are the perfect place for equivalent theories to converge: they are neither too theoretical to avoid idle baggage from either theory, nor too observational to prevent our definition from collapsing into mere empirical equivalence; they are just right.

Since, as said above, the first two steps are captured in the equations of motion, agreeing on step 1 entails that the two theories would generate the same EOMs for the same natural systems.

> *Dynamical Equivalence:* Two theories are dynamically equivalent iff they generate the same equations of motion for every natural system to which they both apply.

To emphasize, I am proposing dynamical equivalence as a form of *theoretical equivalence*, which means that it must be stronger than mere empirical equivalence. And it is, because as said above, an equation of motion is not a "mere observational statement".

The different "theories" of classical mechanics are all dynamically equivalent, for each pair of "theories" generates the same EOMs *whenever both theories apply*, but they do not apply to all of the same systems: each "theory" advises us to encode the transition rules in different functions, which allows that one theory have exclusive applications that the other theories' compactification rules do not permit.

We are now ready to review the recent debates on the (in)equivalence of the different theories of classical mechanics. First off, let us quickly review the existing definitions of

theoretical equivalence in the literature upon which this debate is largely founded.

2.3 Traditional Definitions of Theoretical Equivalence and their Discontents

In this section, I would like to quickly survey the existing definitions of theoretical equivalence in the descriptivist literature and point out some general problems with them.

2.3.1 Traditional accounts

2.3.1.1 Definitional equivalence

One of the oldest and most well-known criteria of theoretical equivalence is *definitional equivalence*, different versions of which have been attributed to Quine and Glymour. The idea behind this notion is straightforward: two theories are equivalent if they are "intertranslatable", i.e. if what each says about the world could just as well be derived from the other theory through new terms defined in the language of the latter. But unlike Quine (1975), who required a direct translation of one theory into the other's language, Glymour (1970, 2013) required that there be a *third language* into which both theories can be translated. As Barrett and Halvorson (2016a) have argued, Glymour's version is less stringent and thus more plausible. Here is one formulation of it due to Barrett and Halvorson:

> *Definitional equivalence:* Let T_1 be a Σ_1-theory and T_2 be a Σ_2-theory [where Σ_1 and Σ_2 are languages or "signatures"]. T_1 and T_2 are *definitionally equivalent* if there is a definitional extension T_1^+ of T_1 to the signature $\Sigma_1 \cup \Sigma_2$ and a definitional extension T_2^+ of T_2 to the signature $\Sigma_1 \cup \Sigma_2$ such that T_1^+ and T_2^+ are logically equivalent. (Barrett and Halvorson 2016a, 470)

The so-called "common definitional extension" T^+ is subject to two conditions: it must say *everything* that either theory says and *nothing* that either one does not say. Other expositions of the idea of definitional equivalence as well as purported counterexamples to it can be found in Halvorson 2012, Barrett 2017, Barrett and Halvorson 2016b, McSweeney

2017, Weatherall 2016a, and Teh and Tsementzis 2017.[5]

2.3.1.2 Structural equivalence

Definitional equivalence is clearly couched in syntactic terms: it requires the specification of languages ("signatures") and definitions to demonstrate the equivalence of theories. Such a definition would not be of much help for an adherent of the semantic view of theories. Rather, a semanticist would presumably seek to establish some sort of equivalence among the *models* or *structures* of the two theories. In light of this, several notions of *structural equivalence* have been proposed. They look roughly something like the following:

> *Structural equivalence:* $\mathcal{M}$ is the same theory as $\mathcal{M}'$, just in case there is a bijection $F : \mathcal{M} \to \mathcal{M}'$ such that each model $m \in \mathcal{M}$ is isomorphic to its paired model $F(m) \in \mathcal{M}'$. (Halvorson 2012, 190)

Halvorson 2012 and Barrett 2017 provide further discussion of and counterexamples to structural equivalence.

2.3.1.3 Categorical equivalence

Critics of structural equivalence claim that it is too permissive, in the sense that it is "too easy" to establish such isomorphic relations among classes of models that intuitively belong to different theories. A more recent shot at defining equivalence attempts to remedy this by loading additional "global" structure (i.e. relations *among* the different models of the same theory) to the class of models, and requiring that this global structure also be preserved. This is the notion of *categorical equivalence*, which was popularized by Halvorson in recent years and has already gained quite a bit of traction. The definition sounds simple enough:

> Theories T_1 and T_2 are categorically equivalent if there is an equivalence between the category of models of T_1 and the category of models of T_2 that preserves the empirical content of the theories. (Barrett 2017, 8)

[5] Barrett and Halvorson 2016b point out that definitional equivalence only allows *predicates* to be translated into each other and therefore fails to obtain when the two signatures Σ_1 and Σ_2 use different *sorts* (e.g. Euclidean geometry couched in terms of points vs. lines). They propose a generalization of definitional equivalence called *Morita equivalence* which allows for sorts to be translated into each other as well.

But there is actually quite a bit of technicality here. "Equivalence" here refers to a technical notion in category theory. As such, a full understanding of the significance of the above definition requires knowledge of category theory, which I shall not venture into here. But roughly speaking, a category is a structured class of objects which, unlike a simple class, has "arrows" defined between the objects. The arrows indicate some notion of identity or interchangeability among the objects and therefore give the class additional structure. An equivalence is a *functor* between two categories that is *full*, *faithful*, and *essentially surjective* (see ibid for definitions of these terms).

The main point to take away is that categorical equivalence is less permissive than structural equivalence because the equivalence functor must preserve the global structures (the arrows) that define relationships *among* the different models as well as establishing an isomorphic bijection between the models (also cf. Curiel 2014, 294 for remarks on the importance of global structure).[6] Note that categorical equivalence has claims to both syntactic and semantic traditions (Nguyen 2017, 985). See also Halvorson 2016, Barrett and Halvorson 2016b, Weatherall 2016a; 2016b for further discussion.

2.3.2 Problems with traditional accounts of equivalence

As we have seen, there are three major definitions of theoretical equivalence (one syntactic, one semantic, and one hybrid) currently in circulation among philosophers of science. Each definition has been criticized for being either too permissive or too strict or both. I shall not delve into these counterexamples here. Instead I would like to point out a few general problems with the definitions presented above.[7]

[6]One could define an even stricter criterion than categorical equivalence by requiring that the two categories of models be *isomorphic* rather than merely *equivalent*. However, Weatherall (2016a, 1081) argues that such a criterion is too strict, for it judges the Faraday tensor formulation of electromagnetism to be inequivalent to the vector potential formulation, which Weatherall seems to consider as a test case for any criterion of equivalence.

[7]There is another general objection to these defintions that I will not go into: these notions of equivalence are purely formal and thus ignore the fact that a formal system is not a theory until it is interpreted in some way. Thus, the above notions of equivalence rule out the possibility that the same formal system may be interpreted in a multitude of ways and be equivalent to another formal system under some but not all of those interpretations. This objection has recently been raised against formal definition of equivalence in general (see Coffey 2014; Nguyen 2017).

2.3.2.1 Lack of distinction between essential and inessential content

First of all, as I mentioned at the beginning of the chapter, a reasonable criterion of equivalence should allow for the possibility that theories contain superfluous baggage that plays no role in their success, aka *idle posits*. Under such circumstances it would seem unnecessary to require that the two theories say all of the same things, seeing as much of what they say is mere embellishment or scaffolding. Yet this is precisely what all of the mainstream definitions of equivalence seem to require.

The lack of attention to idle posits in the mainstream criteria is not simply an oversight. In fact, McSweeney (2017) argues that any criterion of equivalence *must* require that the two theories' contents match *in their entirety*. Allow me to produce a lucidly synoptic paragraph to represent McSweeney's argument:

> [I]f we don't know that T and T' are fundamentally equivalent, it must be that we don't know exactly what their fundamental commitments are. And if this is the case, we can't be justified in *removing* structure from T and T' in order to evaluate them for equivalence, since we don't know which bits of them are metaphysically committing and which are not. Nor can we be justified in restating them in different terms, for doing so would require us to know what that restatement needed to preserve and what it did not. What we *are* justified in doing is checking to see whether we can extend them into a single theory that doesn't remove anything from either one. (McSweeney 2017, 275)

Thus, McSweeney argues that any account of "core equivalence" that allows for non-essential content variation must be a cheat. But the argument crucially hinges on the conditional at the beginning of this paragraph, and this conditional strikes me as puzzling. Consider the following two abilities:

> a) the ability to delineate the "metaphysically committing" (i.e. working) parts of a theory from its superfluous (i.e. idle) parts;
> b) the ability to evaluate the equivalence of the theory with any other theory that is also stripped of its unnecessary content.

McSweeney's argument seems to hold a tacit premise that (a) automatically entails (b). But what is the justification for this premise? Perhaps there is an assumption that the unnecessary parts can be peeled off and the gem of the theory displayed only in *one* particular language, and therefore that all polished theories are in the same language, and thus that if they are equivalent, they will be manifestly and trivially so. This assumption seems

unsupported, however. As we have seen, it is possible to distinguish the theory's prescriptive-dynamical content from its descriptive-ontological embellishments or scaffolding, for instance, without automatically settling the question of equivalence.[8] For instance, it is possible to hold all of the following propositions as true without contradiction:

- The essential content of Newtonian mechanics is assigning states of the form $(x, \dot{x})$ and encoding the transition rules in a vectorial quantity proportional to acceleration.
- The essential content of Lagrangian mechanics is assigning states of the form $(x, \dot{x})$ and encoding the transition rules in a scalar quantity typically but not always equal to $T-U$, where T is kinetic energy and U potential energy.
- It is not clear whether Newtonian mechanics and Lagrangian mechanics are equivalent (until one examines when and whether forces can be converted to Lagrangians and vice versa).

In other words, just because we have isolated the working posits of the theories as their state assignment and compactification rules does not mean that we know the two compactification rules to be equivalent. In other cases the state assignments themselves may not be manifestly equivalent. The Schrödinger and Heisenberg formulations of quantum mechanics provide a well-known example of two state-spaces that were secretly equivalent. This secret would certainly not reveal itself automatically upon the declaration that the the "metaphysically committing" parts of both theories are their state-space structures and nothing more. And as we shall see in a moment, many have doubted whether the Lagrangian states of $(x, \dot{x})$ are equivalent to the Hamiltonian (x, p). Once again, this doubt is fully consistent with the idea that the essential core of both theories is their state-space structure, which many in this debate seem to tacitly assume. In short, one can isolate the working posits of two theories without thereby having settled the question of their equivalence automatically.[9]

[8] McSweeney's argument has resonances with another familiar objection to the idea of working posits raised by Stanford (2003) in response to Kitcher's original introduction of the concept. The objection is that one can never differentiate the working posits from the idle ones *except in retrospect*, i.e. after one has already established what was right and what was wrong about the theory. As I shall argue in Chapter 3, this is also remedied in the prescriptive-dynamical account for the same reason as above.

[9] A weaker, more plausible version of McSweeney's claim would be that one cannot expect to commit to an essential / non-essential distinction in theoretical content without thereby having made some commitments about what theoretical equivalence amounts to. This is of course the thought with which we began the chapter, and which brings me to my last objection to traditional notions of equivalence.

2.3.2.2 Lack of applicability to physical theories in practice

The next general problem with the mainstream definitions of equivalence is that these definitions are couched in frameworks that have questionable application to the practice of science. Hardly any major physical theory has been convincingly formulated in terms of a first-order language with predicate logic, model theory, or category theory, which creates a serious disconnect between philosophy of science and the science it is meant to be a philosophy of. This problem is in fact occasionally recognized in passing by the very authors who continue to use such frameworks (see Barrett 2017, 7; Weatherall 2016a, 1080), only to be dismissed in the hopes that the conclusions remain general nonetheless.

It is not surprising that these accounts of equivalence have found little success in being applied to the practiced theories of physics. After all, the above-mentioned accounts of equivalence, whether syntactic or semantic or hybrid, all operate under the assumption that the theory is a fixed set. This requires, among other things, that two theories must share all of the same models, i.e. say the same things about every natural system that they apply to, in order to be equivalent. Basically, all mainstream definitions of theoretical equivalence are versions of compact equivalence, which as argued above is irrelevant to the practice of physics.

Once one realizes that the transition rules are in practice open-ended, any notion of a bijection between two theories that share all of their consequences / models becomes irrelevant. There cannot be a guarantee that all of the future models of the two theories will overlap, because one cannot predict future LEMPs due to inexhaustibility. Theories can only be (in)equivalent with regard to their fixed parts, which is the state assignments. As such, any criterion of equivalence that requires a one-to-one match-up or translation of theorems or models of the two theories is going to be inapplicable to physical theories (though perhaps useful for mathematical theories). As a result, the descriptivist must either find artificial reasons for restricting herself to a subset of the two theories' models which does fully overlap, or else admit that she cannot make sense of the ubiquitous claims of equivalence by physicists.

Since the issue of restrictions on LEMPs is clearly demonstrated in discussions of equiv-

alence in classical mechanics, I will suffice it to the above for now. Let us see how all of this plays out in the context of the recent debates on classical mechanics.

2.4 A Critical Review of Recent Debates on Theoretical Equivalence in Classical Mechanics

2.4.1 Applications of traditional accounts of equivalence to classical mechanics

The recent resurgence of interest in theoretical equivalence was sparked by a paper by North (2009). North seems to be implicitly working with structural equivalence as the criterion of theoretical equivalence (cf. Barrett 2017, 3). She begins by suggesting that physics tells us about the structure of the world, and goes on to ask how this structure is to be identified given a physical theory, especially when there are genuinely different formulations of the theory, such as one finds in the Heisenberg vs. Schrödinger formulations of quantum theory or in Lagrangian and Hamiltonian mechanics. After several examples, North concludes that structure in general must be modeled after *geometric structure*, which she defines as whatever is invariant under coordinate transformations (ibid, p. 61-2). A vector as a mathematical object is invariant under coordinate transformations but its components are not. Therefore, the vector is part of the "structure" whereas its components are not. Thus, according to North, one finds the structure of the world as described by the theory by transforming coordinate- and frame-dependent descriptions of the theory into each other and identifying the invariant objects. (As a result, North adds yet another definition to the menu of definitions of "structure", ostensibly different from the "mathematical equations" view, Ramsification, relations, and relations of relations. See Chapter 3 for a menu of such definitions.)

The gist of North's argument is that Hamiltonian and Lagrangian theories have different state-space structures, because the former is a symplectic manifold or a cotangent bundle with just a volume element, while the latter is a tangent bundle which comes with a metric.

> Transformations from one set of Lagrangian coordinates to another ... preserve [the Riemannian line element] in the form of a quadratic differential form of the $\dot{q}$'s. This allows us

> to say that the structure of the Lagrangian statespace is that of a Riemannian manifold, with a Riemannian metric defined on it.
> The statespace of Hamiltonian mechanics – phase space, the cotangent bundle – is different. Phase space has symplectic structure; it does not have metric structure. Symplectic structure comes with, or determines, a volume element, but not a distance measure. (ibid, p. 73)

North appeals to a sort of "structural Ockham's razor" to conclude that Hamiltonian mechanics should be preferred, but that particular conclusion does not concern us here as our focus is the question of equivalence itself.

Curiel (2014), seemingly working with a criterion along the lines of categorical equivalence, disagrees with North on the mathematical details and on which of the two theories must be preferred, but the two authors agree on the inequivalence of the two theories. In fact, Curiel draws a conclusion far stronger than North's regarding this lack of equivalence. The gist of Curiel's argument is that classical states are characterized by two quantities: a *configurative* one such as q and a *velocital* one such as $\dot{q}$. While one cannot through mere measurement find which measurable quantities are configurative and which ones are velocital, Curiel's argument goes, we know through experience that the configurative quantities have a "fixed" equation of motion, such as $\dot{x} = v$, which is to say that their equation of motion does not depend on the system or the type of interactions it is involved in. Velocital quantities, on the other hand, have an "unfixed" equation of motion such as $\dot{v} = \frac{F}{m}$, which varies from system to system. It is this difference in the fixity of the equations of motion that distinguishes configurative variables from velocital ones. Now, the crux of Curiel's argument is that $\dot{x} = v$ is an unchangeable *kinematical constraint* in Lagrangian mechanics, whereas the same formula is not even required by Hamiltonian mechanics and can be violated in many examples. Therefore, Curiel goes on to conclude, Lagrangian mechanics describes classical systems whereas the Hamiltonian theory does not.

> One way to see this is that one has no way [in Hamiltonian mechanics] to formulate the kinematical constraints appropriate to Lagrangian mechanics [e.g. $\dot{x} = v$] for one has no almost-tangent structure, but those kinematical constraints are the natural ones for abstract classical systems.
> In sum, then, the family of dynamical evolutions Hamiltonian mechanics admits for systems it represents is not isomorphic to the family of kinematical vector fields of an abstract classical system [which *is* isomorphic to the Lagrangian structure]. (ibid, p. 298)

Of course if one were to stipulate that $p = mv$, Hamilton's first equation would entail $\dot{x} = v$. But the problem is that there is no basis in the Hamiltonian structure for assuming that $p = mv$.

What about Legendre transformation? Doesn't it convert the Lagrangian and Hamiltonian models to each other? First of all, Legendre transformations only work on a subset of all possible Lagrangians (and likewise the reverse Legendre transformations only apply to a subset of possible Hamiltonians). Most natural systems can be assigned both a Lagrangian and a Hamiltonian, although natural systems with nonconservative forces, which have no Hamiltonian model, can sometimes be treated with inhomogeneous Lagrange equations. This and the other facts mentioned above lead Curiel to the conclusion that the Legendre transformations take Lagrangian models to Hamiltonian models that quantitatively agree with them but that *mean something else entirely*:

> Thus, even when Lagrangian and Hamiltonian mechanics provide individual models of the same physical system, no structure in the individual model of the one is isomorphic to any in the other in a way that has real, non-trivial physical significance... . The only way to deny this conclusion is to claim that sameness of solutions to equations of motion by itself – mere sameness in brute description of motion – ensures sameness of physical significance, but that is nothing more than the most naive form of verificationism.[10] (ibid, p. 314)

A flurry of inquiry ensued the exchange between North and Curiel. Here are the highlights. Swanson and Halvorson (2012) take issue with many technical aspects of North's analysis. As it turns out, comparing structures is harder than North had assumed: contrary to North's analysis, Swanson and Halvorson point out that symplectic manifolds have symplectic forms as well as orientations whereas metric spaces do not have orientation. Therefore, they argue, there is no straightforward sense in which Hamiltonian mechanics has "less structure" than Lagrangian mechanics. We will return to some other complications pointed out by the authors below.

Barrett (2014) also takes issue with many of North's assertions. Barrett argues, contrary to North, that the Lagrangian state-space is not necessarily always endowed with a metric structure, and that the Hamiltonian phase space often has a metric structure (especially when the Lagrangian counterpart does). Barrett also echoes Swanson and Halvorson's ob-

[10]Note Curiel's talk of sameness of *solutions* to equations of motion, which is empirical equivalence, rather than sameness of equations of motion, which is my notion of dynamical equivalence.

jection that due to the existence of orientation on symplectic manifolds, such structures are "incomparable" to tangent bundle structures. These are important technical disagreements, but they do not change the verdict that the two theories are inequivalent because their state-spaces are non-isomorphic.

In a more recent paper, Barrett (2017) argues that whether Lagrangian and Hamiltonian mechanics are (categorically) equivalent depends on many factors. First of all, Barrett limits his inquiry to *hyperregular* models of both theories (Barrett 2017, 10-13). These are essentially models that are amenable to Legendre transformation (Abraham and Marsden 1987, 218 ff.). This restriction does not automatically settle the question of equivalence, because as we saw part of the debate concerns the fact that even when a bijective Legendre transform is available, this transformation is not structure-preserving (Curiel 2014, 308 ff.). Nevertheless, the restriction to Legendre-amenable models is a substantial assumption as far as equivalence goes. Consider for example the fact that much of Curiel's arguments regarding the inequivalence of the two theories hinges on the fact that not all models have well-defined Legendre transforms: "the Legendre transform does not always map all Lagrangian vector fields to Hamiltonian ones" (ibid, 309) and "not all Hamiltonian vector fields map to second-order [i.e. Lagrangian] vector fields under the inverse-Legendre transform." (ibid, 310) Given the dialectical stakes, it is somewhat puzzling that Barrett gives no justification for the restriction to hyperregular models.

In any case, even the restriction to hyperregular models is not enough to do the trick. Having argued for categorical equivalence as the superior criterion of equivalence, Barrett shows that if one defines Hamiltonian mechanics on a symplectic manifold, then it is not categorically equivalent to Lagrangian mechanics, but if one defines it on a cotangent bundle, then it is so (Barrett 2017, 14 ff.). To sum up, Barrett shows that hyperregular Hamiltonian mechanics defined on a cotangent bundle is equivalent to Lagrangian mechanics. Otherwise the two theories are inequivalent.

Teh and Tsementzis (2017) reach a similar conclusion using *definitional equivalence* as their criterion. Having restricted the case to hyperregular models only, Teh and Tsementzis proceed to show that Lagrangian theory as formulated on a tangent bundle TQ of the configuration space and Hamiltonian theory as formulated on a cotangent bundle T^*Q, are

definitionally equivalent if we "lift" both theories to the cotangent bundle of the tangent and cotangent bundles (T^*TQ and T^*T^*Q). Following a clever method devised by Tulczyjew (1974, 1977), the authors then show that the tangent bundle of the tangent and cotangent bundles TT^*Q can serve as the "third language" in which to build the common definitional extension of the two "lifted" theories. Thus, Teh and Tsementzis reach the same verdict as Barrett albeit using definitional equivalence rather than categorical equivalence: if Hamiltonian mechanics were hyperregular and formulated on a cotangent bundle (as opposed to a symplectic manifold), then it is equivalent to Lagrangian mechanics.

2.4.2 Problems with recent debates on theoretical equivalence

There are two main problems with this debate, both having to do with the two-tier character of theories that has been my main focus: 1) some parts of the debate violate the fixity of state assignment rules, and 2) other parts of the debate violate the open-endedness of transition rules. Let us examine these problems in turn.

2.4.2.1 The problem of open-ended state assignments: A Tale of Two Hamiltonians

If what I have been arguing is correct, then physical theories have fixed state assignments and open-ended transition rules. My emphasis has been mostly on the latter fact, as it is typically the neglected one; but the former fact is of course of equal importance. Indeed, some of the deepest problems with the debate on theoretical equivalence in classical physics can be traced back to a violation of the requirement that a theory's state assignments be fixed before the transition rules are read off. Allow me to explain.

From a traditional, descriptive-ontological point of view, it appears as though one can formulate a theory without any direct assignment of the dynamical states to experimental procedures. To be sure, the theory must connect to experience *at some point* for it to have empirical import, but traditionally this connection need not be at the level of state assignments: it can be much further down the chain of inference. This stems from the fact that in traditional accounts, "the theory", which is a fixed set of sentences or models, comes

equipped with a state-space *as well as a list of specific transition rules* (e.g. Hamiltonians) that are somehow handed down to us in a pre-ordained manner, as if in a physics exercise. One then puts the states and the transition rules together to infer a trajectory in state-space, which can then be projected on the configuration space, giving $x(t)$, say, as the final prediction of the theory.

State-space
Transition rules
Initial conditions
———————
Trajectory

Once the trajectory is given as a mathematical function, all one needs is some set of "correspondence rules"[11] (as well as auxiliaries, as the case may be) to tell us how to measure the parameters characterizing the trajectory (e.g. x and t), and we are good to go:

Trajectory
Correspondence rules
(Auxiliary assumptions)
———————
Empirical prediction

Descriptivist accounts therefore take this final product, i.e. the physical trajectory, to be the only necessary point of contact between theory and experience. In other words, in traditional accounts, the theory need not connect to measurement *until after the transition rules are plugged in and a trajectory is inferred.*

The situation is fundamentally different in the prescriptive-dynamical view: the "correspondence rules" must be provided for all components of the state assignment *before* transition rules can be recorded to begin with. In other words, the theory's state assignments must have a fixed and unambiguous connection to experimental procedures, independently of what the local transition rules may turn out to be. This is of course due to the fact that

[11] I use "correspondence rules" very loosely in the sense of whatever unspecified (and likely complex) information is needed to relate theoretical quantities to measurement results. My arguments do not depend on the specifics of how these rules work, and in particular whether correspondence rules establish direct or indirect links with observation and whether theoretical terms should be considered "meaningless" prior to the application of the correspondence rules. For discussion see Hempel (1958; 1966), Carnap (1953; 1966), Nagel (1961), Hesse (1965), Schaffner (1969), Halvorson (2016)).

transition rules are "reverse engineered", by first assigning the prescribed states and then following them along as they evolve. This means that the theory *must* connect to experiment at the level of state assignment rules, long before any prediction is made. One must be able to prepare or detect the system in some of the proposed states or else there will be no transition rules. So from a prescriptive standpoint we have:

State assignment rules
Correspondence rules
Empirical data
———————-
Transition rules

One then puts this together with the initial conditions to get a unique trajectory:

State assignment rules
Transition rules
Initial conditions
———————
Trajectory

The next steps are the same as in traditional accounts.

I shall refer to the statement that the theory connects to experience at the level of state assignments as *Early Point of Contact (EPOC):*

> EPOC: The state assignments of the theory must be *experimentally fixed* prior to the construction of LEMPs.

EPOC is necessary for a theory to be *productive*. By "productive" I mean that the theory allows the scientist to generate original LEMPs through direct engagement with the system *upon the discovery of a new system or local context of interaction* for which no LEMP has been recorded yet.

> *Productivity:* A theory is productive iff it enables the scientist to produce potentially novel transition rules for new systems and contexts.

In order for this to happen, it is not enough that the scientist treat the set of transition rules as open-ended; it is also necessary (and sufficient) that knowledge of transition rules

not be required for connecting the state assignments to experimental procedures, i.e. that the theory satisfy EPOC.

Bohr's atomic theory provides a great illustration of a productive theory. The state assignments of this theory are given by the Bohr-Sommerfeld quantization condition, while the transition amplitudes are encoded in the Correspondence Principle (see Kaveh 2014). Experimentally, the state assignment rules can be inferred from the spacing between different spectral lines, whereas the transition amplitudes must be read off the brightness of the lines. Consequently, Bohr was able to devise experimentally unambiguous state assignment rules that can be applied independently of the transition rules. This is important, because without the assignment of Bohr's stationary states, there is no link between the intensities of spectral lines and any notion of transition amplitude. The connection is only established after one has decided on a way to assign initial and final orbit labels to each spectral line, which requires state assignments. Bohr's theory therefore allows one to reverse engineer an appropriate correspondence principle to encode the amplitudes, just as a proper productive theory should.

By contrast, if a theory violates EPOC, and hence requires *both* a state-space *and* a LEMP to be provided before it can connect to experience, then the theory is infertile: there is no way to read off the local transition rules for a new system or context of interaction, because by assumption the state assignment rules of the theory depend on what the transition rules are. This means that the theory must borrow both its state assignment rules and its transition rules from somewhere else. Such theories must therefore purloin their state assignments and LEMPs from productive theories if they are to have any empirical applications. I shall refer to such theories as *unproductive theories*.

An unproductive theory would have no application without another theory to piggy-back on, for one would not know how to begin applying the theory. We cannot have transition rules unless we have states, and if the states are open-ended, there is nothing to guide what one should keep track of. But if one could find another theory with fixed state assignments to piggy-back on, one could have a "theory" of some kind by constructing a "translation manual" between the two formalisms that fixes both state assignments and transition rules.

The most prominent example of an unproductive theory with open-ended state assign-

ments is Hamiltonian mechanics when it is considered as parasitic on Lagrangian mechanics. Textbooks of physics often treat Hamiltonian mechanics this way by presenting it as always being the result of a Legendre transformation – a mathematical technique that connects the Lagrangian formalism to that of Hamiltonian mechanics. One begins with a Lagrangian L and assigns $p_i = \frac{\partial L}{\partial \dot{q}_i}$ and $H = \Sigma_i p_i \dot{q}_i - L$. The momentum p often turns out to be mv or a similar expression that is linear in generalized coordinates (something like $p_i = \Sigma \lambda_i^j \dot{q}_j$). But there is no guarantee. The momentum p need not be linear in velocities and might turn out to be any number of things depending on what L looks like. Nevertheless, the resulting Hamiltonian model with (x, p) as the states and H as the encoder of the missing ingredient for transition rules will always yield the same equations of motion as the Lagrangian with $(x, \dot{x})$ as states and L as the missing ingredient for transition rules. Hamiltonian mechanics as presented in these textbooks is therefore parasitic on Lagrangian mechanics, because without knowing L, one would not know what to assign p to. Every time we stumble upon a new natural system or a new context of interaction (say, electromagnetic fields), we need to consult Lagrangian mechanics first to know what the state assignments and LEMPs of Hamiltonian mechanics are going to be. This theory does not come with a manual of what types of states to assign under what types of circumstances. Lagrangian mechanics tells it what to do.

Now, to be clear, there is nothing special about Hamiltonian mechanics that forces it to be a parasitic theory: *both* theories can be autonomous if they fix their state assignments independently. And to emphasize, they must fix their state assignments *experimentally*: Lagrangian mechanics must tell us what experimental procedures measure x and $\dot{x}$, and Hamiltonian mechanics must do the same for x and p. The question of *how* precisely this is done in the case of a productive theory and how precisely it fails in the case of an unproductive one can get rather intricate. I have provided more technical details regarding experimental state assignments as well as an example of a system for which the Hamiltonian model is unproductively obtained in Appendix A below.

To sum up, there are at least two things that might be called "Hamiltonian mechanics":

i) an *autonomous* and hence *productive* theory with fixed state assignments but narrower application;

ii) a *parasitic* and hence *unproductive* theory with open-ended state assignments but wider application.

A scientist working with the autonomous theory begins by assigning states of the form (q, p) to the system, where q and p are attached to known experimental procedures that determine their value, and proceeds to extract transition rules from the data and compactify them in the quantity H. A scientist applying the parasitic Hamiltonian formalism, on the other hand, begins with assigning Lagrangian states of $(x, \dot{x})$, goes on to record the local empirical transition rules, encodes them in a Lagrangian, and then proceeds to assign $p = \frac{\partial L}{\partial \dot{q}}$ and compute the Hamiltonian through a Legendre transformation. In the autonomous theory, the momentum p must be assigned to a fixed measurable quantity *for all systems and context to which the states apply*. As a parasite to Lagrangian mechanics, however, p is assigned to something different every time, and is therefore undetermined until one is given the transition rules as encoded in a Lagrangian. The autonomous Hamiltonian satisfies EPOC; the parasitic Hamiltonian does not. The former is productive, the latter unproductive.

Now, what do our authors mean by "Hamiltonian mechanics"? As I will now argue, the authors in this debate have been working with a "Hamiltonian theory" that is presumably an autonomous theory, but one that also has open-ended state assignments. As I have argued, an autonomous physical theory cannot have open-ended state assignments. Thus, if I am right, all of the foregoing debates regarding theoretical equivalence in classical mechanics should be considered irrelevant to empirical science. Let me justify this accusation.

Take Curiel for instance. What Curiel refers to as "Hamiltonian mechanics" seems to be an autonomous theory. For example, he defines x and p entirely abstractly, without connection to any physical quantity (Curiel 2014, 296 ff.). He argues that Hamiltonian systems can violate $\dot{x} = v$ (ibid, 298), and makes use of "funny" examples that have no physical relevance (ibid, 305-6). He also claims that $p = mv$ is an "ad hoc" condition in Hamiltonian mechanics, whereas for a parasitic Hamiltonian system that piggy-backs on Lagrangians, the relation $p = mv$ is merely the result of the Legendre transformation of the usual kinetic term. In line with this, Curiel argues that Hamiltonian models that apply to classical systems are "a small class" of all Hamiltonians (Curiel 2014, 304). These facts indicate that the Hamiltonian theory Curiel has in mind is *not* the one that is obtained by

applying a Legendre transformation to a Lagrangian, but one that starts with its own axioms and geometrical structure, and allows Hamiltonians that have no Lagrangian representation.

The second fact, namely that what Curiel has in mind by "Hamiltonian mechanics" has open-ended state assignment rules, can be inferred from the following observations. Curiel claims that "one ... loses the capacity to identify configuration space in the Hamiltonian formulation" (Curiel 2014, 298). This means that the Hamiltonian q is not necessarily position. But neither is it assigned to any fixed quantity other than position, for Curiel argues that one can "mix up" or even "swap" q and p at will. Moreover, we are told that one must "impose" relations of the form $p_i = \Sigma\lambda_i^j \dot{q}_j$ "by hand", which implies that without this intervention, p is not assigned to any experimental procedure: if both p and q were independently measurable, then relations of the form $p = f(\dot{q})$ would be *empirical* facts, not conditions imposed "by hand". Finally, Curiel does not shy away from examples that involve unphysical momenta (ibid, 314). These facts indicate that in Curiel's "Hamiltonian theory", q and p are not assigned to any fixed experimental procedure.

Therefore, what Curiel has in mind by "Hamiltonian mechanics" is a theory that is a) autonomous of Lagrangian mechanics, and b) has no fixed state assignment rules. This is not a legitimate empirical theory and as such, Curiel's analysis is fundamentally comparing apples to oranges.

The situation is no better with the other authors. First of all, as we saw, both Barrett and Teh and Tsementzis work with a Hamiltonian theory that is restricted to its hyperregular models. The fact that one needs to restrict this theory before it can be equivalent to Lagrangian mechanics clearly implies that this "Hamiltonian theory" is not parasitic on Lagrangian mechanics but rather an autonomous theory. Secondly, hyperregular models are those that are amenable to Legendre transformations. If the authors intend to include all hyperregular models, then as we saw above, this includes perfectly physical Lagrangians whose Legendre transform results in odd momentum assignments that do not correspond to the quantity transferred in collision. Therefore, even though this theory is meant to be autonomous, it has no experimentally fixed state assignment rules, which once again means we are dealing with a non-physical theory.

So, the entire debate regarding the equivalence of Lagrangian and Hamiltonian mechanics

rests on equivocation on what "Hamiltonian mechanics" refers to.

2.4.2.2 The problem of restricted transition rules

As we saw, those authors who argue that Lagrangian and Hamiltonian mechanics are descriptively equivalent reach this conclusion by restricting our attention to a subset of each theory's models, typically that of the hyperregular models on a cotangent bundle. Why this restriction? First of all, note that one could *not* justify such a restriction by pointing out that physicists *typically* work with models that are amenable to Legendre transformation (which is the property that picks out hyperregular models). The question of interest to us is whether the two theories are equivalent, and this question is independent of which models the physicist chooses to *focus* on. "Hyperregular Lagrangian" and "hyperregular Hamiltonian" do not designate theories of their own. Besides, what about those times when physicists do use models of Lagrangian mechanics that are not hyperregular? Whatever else one might say about this, it is clear that physicists are not switching to a different theory in doing so. Therefore, there is nothing natural or physical about the restriction to hyperregular models.

As I argued above, such arbitrary restrictions are to be expected if one is trying to make sense of physicists' unanimous and unrelenting insistence on the equivalence of Lagrangian and Hamiltonian theories *in a descriptivist framework*. This is because the two theories have non-overlapping models, whereas definitional equivalence, structural equivalence, and categorical equivalence all require fully overlapping models. Consequently, the physicist must be judged flatly wrong, if not idiotic, for making such careless pronouncements unless one finds a way to restrict one's attention to the overlapping subset of models only.

Now, the problem is that restrictions to a subset of models amounts to restrictions on permissible *LEMPs*: the restrictions to hyperregular models and to cotangent bundles are both restrictions on the types of Lagrangian / Hamiltonian that one is willing to allow. This is problematic given the open-endedness of LEMPs. As I argued in Chapter 1, it is important for physical theories to satisfy ABUNDANCE, i.e. not to put a priori restrictions on the types of transition rules that are allowed. If theoretical equivalence can only be shown by fixing the LEMPs, then equivalence of *physical theories* can never be shown, for

mathematical theories cease to be physical theories as soon as they are fixed.

Restricting the LEMPs serves as a sneaky way of ruling out those pesky non-overlapping models that make trouble for equivalence claims. But that is not all. Restrictions on LEMPs cause even deeper distortions in the project pursued by the authors above and threaten to undermine traditional notions of equivalence as well as structuralist approaches more generally. This is because the structure of state-space itself depends on the choice of LEMP. As soon as one specifies a Lagrangian or Hamiltonian function on the manifold, the "structure" of the state-space is altered. In fact, this has been recognized in the literature, although its catastrophic consequences for the above inquiries have not been acknowledged due to the assumption that theories are fixed. Swanson and Halvorson discuss the dependence of state-space structure on the choice of LEMPs as a criticism of North's argument:

> Part of the difficulty in making broad structural comparisons the way North does, is that neither Hamiltonian nor Lagrangian mechanics purports to be a complete description of the world. A complete physical model is generated only once particular forces and dynamical laws [i.e. LEMPs] are specified, along with how the statespace description of these structures maps onto spacetime. By focusing her attention on the two theories abstractly, and not on particular, fully detailed models, North's analysis only captures the features of a generic model, the structure that every model of Hamiltonian or Lagrangian mechanics has in common. But most models have more structure, and much of this structure appears, on a natural interpretation, to be physically important. (Swanson and Halvorson 2012, 9)

Barrett also mentions the effect of LEMPs on state-space structure in a similar fashion:

> [North] demonstrates that Lagrangian mechanics has metric structure by considering both the statespace T_*Q [a tangent bundle] and the Lagrangian $L : T_*Q \to \mathbb{R}$ defined on that state-space. The argument that Hamiltonian mechanics has symplectic structure, however, only relies on the Hamiltonian state-space T^*Q [a cotangent bundle]. It does not take into account the Hamiltonian $H : T^*Q \to \mathbb{R}$ defined on the state-space.
> When this asymmetry is remedied, one can see that insofar as the Lagrangian state-space is equipped with the [metric] structure g_q, the Hamiltonian state-space is equipped with the [metric] structure g_q^*. (Barrett 2014, 811)

Indeed, as Teh and Tsementzis point out, even the meaning of "isomorphism" itself depends on the choice of Lagrangian (Teh and Tsementzis 2017, 46-47)

Now, as mentioned above, these authors do not consider the dependence of state-space structure on transition rules to be a fatal issue for their inquiry, because the assumption in the literature is that the transition rules are somehow given to us, handed down from the

sky, as it were (see, e.g., how Barrett characterizes a "model" of Hamiltonian or Lagrangian theory to include a specification of a Hamiltonian / Lagrangian, respectively (Barrett 2017, 10, 11)). Under such an assumption, there is no major crisis: all one needs to do is make sure one has a full catalog of all the relevant LEMPs (Lagrangians and Hamiltonians) before comparing structures. This sentiment is expressed by Swanson and Halvorson as follows:

> The interpretive question which motivates the structuralist razor is not, "which of these two theories posits less *common* structure between its models," but rather "which of these two theories posits less *total* structure on the world." It doesn't help to know that the generic Hamiltonian dynamics only require statespace to be a symplectic manifold if additional structure must be posited on a model-by-model basis in order to illuminate how this statespace description relates to structures in spacetime. (Swanson and Halvorson 2012, 9 – emphasis in the original)

But there is no such thing as "total structure" unless the LEMPs are fixed. Since altering the LEMPs alters the structure of the theories, to know the total structure one must first know the complete set of LEMPs. However, as I have been arguing all along, *there is no such thing as a complete set of LEMPs*! The LEMPs are open-ended. The open-endedness of LEMPs implies that the state-space structure itself is open-ended. Consequently, structural comparisons can never settle questions of equivalence: no fixed LEMP, no fixed structure; no fixed structure, no descriptive equivalence. More generally, structuralism itself appears to be chasing after a ghost.

On a more dialectical note, one can see where the source of disagreement in the literature is: since LEMPs are open-ended, the (in)equivalence of theories cannot really be established unless one does one of two things: allow all possible LEMPs, or allow a subset of "well-behaved" LEMPs. North and Curiel do the former and conclude that the two theories are inequivalent, while Barrett and Teh and Tsementzis do the latter and conclude the opposite. As we saw in Chapter 1, both options regarding what LEMPs to allow are problematic (see Ch. 1, Section 6, the Descriptivist's Dilemma). LEMPs can be neither restricted nor assumed to cover all possible transition rules; they must remain open-ended.

Let us wrap up this section by summarizing how things stand in the prescriptive-dynamical view.

> 1- The autonomous Hamiltonian mechanics and Lagrangian mechanics are *dynamically equivalent*. They generate the same equations of motion for any natural systems to which

they both apply, but the Hamiltonian framework will fail to compactify many such EOMs in certain cases that the Lagrangian theory treats just fine.
2- The autonomous Hamiltonian mechanics and Lagrangian mechanics are not *compactly equivalent*, since there are Lagrangian systems that have no well-defined Hamiltonian counterpart.
3- The parasitic Hamiltonian mechanics is "unproductively equivalent" to Lagrangian mechanics, meaning it encodes the same EOMs, but must derive its state assignments from a productive Lagrangian theory.

In the final section below, I would like to present evidence that dynamical equivalence is precisely what a practicing physicist means when they claim that Lagrangian and Hamiltonian mechanics are equivalent.

2.5 Dynamical Equivalence is the Practicing Physicist's Notion of "Equivalence"

I claim that dynamical equivalence as defined above is not only the notion of equivalence that makes sense of scientific practice, but also one that the physicist all but explicitly adheres to. One way to justify the latter claim is to survey physicists on what they mean by theoretical equivalence. But that method is unlikely to produce satisfactory evidence one way or another, as physicists' responses to philosophical questions tend to be largely half-baked, incoherent both within and between individuals, and couched in philosophically sloppy terms. But this should not worry the philosopher, for physics is not done in words, but in actions. More important than what physicists *say* in response to a survey question is how they are *trained* from the ground up to interpret and manipulate symbols on a page, and what significance they are taught to attach to these symbols. Whatever notion of equivalence arises from this early training is likely to carry through the physicist's career. And what better place to look for the fundamental characteristics of this training than the widely used textbooks of foundational physics? Since the recent debates on theoretical equivalence have been focused on classical mechanics, in this section I will present textual evidence for the claim above from some of the most prominent textbooks of classical mechanics.

Recall that dynamical equivalence comes down to the two theories generating the same

equations of motion. I argued that this notion of equivalence is stronger than mere empirical equivalence (agreement on trajectories) and yet allows for the idle parts of equivalent theories to be in disagreement. I will now present textual evidence that this is exactly how physics students are trained to understand equivalence.

Let us begin with Patrick Hamill's *A Student's Guide to Lagrangians and Hamiltonians*, which provides one of the most lucid expositions of the idea. The section in which Newton's second law and Lagrange's equations are introduced is called "Obtaining the equation of motion". Hamill quotes Landau and Lifshitz saying that once coordinates and velocities are given, "the state of the system is completely determined and ... its subsequent motion can, in principle, be calculated", which matches my characterization of dynamical states and their role in physical theories. Hamill then goes on to point out that assigning states of the form $(q, \dot{q})$ is equivalent to writing a second-order differential equation:

> In other words, the equations of motion can be expressed as relation of the type:
>
> $$\ddot{q}_i = \ddot{q}_i(q_1, q_2, ..., q_n; \dot{q}_1, \dot{q}_2, ..., \dot{q}_n; t) \tag{5}$$
>
> (Hamill 2014, 18)

which once again matches my characterization of the essential content of classical mechanics. Hamill then goes on to introduce Newtonian and Lagrangian systems as ways of obtaining the EOM:

> If the masses are constant, an elementary way way to obtain the equation of motion is to use Newton's second law in the form
>
> $$\ddot{\mathbf{r}} = \frac{\mathbf{F}}{m}. \tag{6}$$
>
> Another way to obtain the equation of motion is to use the Lagrangian technique. (Hamill 2014, 19)

The same sentiment is expressed vividly in a footnote: "In general, the Lagrangian is defined to be a function that generates the equation of motion." (ibid) The same idea is also conveyed upon the introduction of the Euler-Lagrange equations: "use it as a tool for obtaining the equation of motion" (ibid, 20). And similarly when Poisson brackets are introduced:

> We have expressed the equations of motion for a mechanical system in a variety of ways, including Newton's second law, the Lagrange equations, and Hamilton's equations. The equations of motion can also be expressed in terms of the Poisson brackets, as we now demonstrate. (Hamill 2014, 119)

In Volume 1 of their *Course of Theoretical Physics* Landau and Lifshitz agree:

> Like Lagrange's equations and the canonical [Hamiltonian] equations, the Hamilton-Jacobi equation is the basis of a general method of integrating the equations of motion. (Landau and Lifshitz 2000 [1960], 148)

One finds the same approach in Goldstein's legendary textbook on classical mechanics:

> [F]or systems where one can define a Lagrangian, ..., we have a very convenient way of setting up the equations of motion. (Goldstein 1980 [1950], 25)

The idea being conveyed is that the value of these different formulations is purely instrumental and one should hold the equations of motion in mind as the ultimate end. This is expressed more explicitly in some moments:

> In view of the difficulties in formulating a variational principle for nonholonomic systems, and the relative ease with which the equations of motion can be obtained directly [i.e. via Newton's method], it is natural to question the usefulness of the variational approach in this case. It is for this reason that discussions of variational principles and their consequences will be confined from here on to holonomic systems in which generalized coordinates are independent. (Goldstein 1980 [1950], 49)

The different formulations are thus distinct for the physicist only insofar as they provide different compactifications:

> We have seen that the Lagrangian and Hamilton's principle together form a compact invariant way of implying the mechanical equations of motion. (Goldstein 1980 [1950], 53)

That is why upon introduction of Hamiltonian mechanics, Goldstein declares that "nothing new is added to the physics involved" (Goldstein 1980 [1950], 339) when one switches from one formulation to another.

In his *Classical Mechanics*, Gregory similarly cashes out the equivalence of Lagrangian and Hamiltonian mechanics in terms of equations of motion:

> Lagrange's equations of motion can be derived from Hamilton's principle, which can therefore be taken as the basic postulate of classical mechanics, instead of Newton's laws. (Gregory 2006, 367)

Also note how Gregory qualifies the above claim a few pages later:

> [H]amilton's principle ... can be regarded as the fundamental postulate of classical mechanics, instead of Newton's laws, *for any mechanical system that has a Lagrangian.* (Gregory 2006, 387 – emphasis added)

From a descriptivist point of view, Gregory's is a contradictory statement: how can two theories be "equivalent" but "only for such and such cases"? Equivalence tends to be a matter of all or nothing for the descriptivist. However, as we have seen, the prescriptivist vantage point makes perfect sense of this: the two theories are equivalent insofar as they generate the same EOMs *whenever both compactification rules are applicable*, i.e. they agree on the state assignments (dynamical equivalence), which allows for the possibility of models that are not shared between the two theories.

The idea of dynamical equivalence is also expressed quite clearly in Morin's *Introduction to Classical Mechanics*:

> Consider the system of a mass on the end of a spring. We can analyze this, of course, by using $F = ma$ to write down $m\ddot{x} = -kx$. ... We can, however, figure things out by using another method which doesn't explicitly use $F = ma$. In many (in fact, probably most) physical situations, this new method is far superior to using $F = ma$.
> For the problem at hand, [the Euler-Lagrange equation] gives $m\ddot{x} = -kx$ which is exactly the result obtained by using F = ma. An equation such as [the above] is called an *equation of motion.* (Morin 2007, 218-219 – emphasis in the original)

And a few pages later, Morin declares:

> At this point it seems to be personal preference, and all academic, whether you use the Lagrangian method or the $F = ma$ method. The two methods produce the same equations [of motion]. However, in problems involving more than one variable, it usually turns out to be much easier to write down T and V, as opposed to writing down all the forces. (Morin 2007, 221)

Taylor similarly cashes out the equivalence of Newtonian and Lagrangian theories in terms of their capacity to generate equations of motion and their differences in terms of compactification:

> The last three examples are sufficiently complex that solution using the Newtonian approach requires considerable ingenuity; by contrast, the Lagrangian approach *lets us write down the equations of motion almost without thinking.*
> The Lagrangian formalism always (or nearly always) gives a straightforward means of writing down the equations of motion. (Taylor 2005, 254 – emphasis added)

And similarly for Hamiltonian mechanics:

> [I]t is a wonderful property of Hamilton's approach (like Lagrange's) that it provides an almost infallible way to find the equations of motion. (Taylor 2005, 528)
> [W]e see that the Hamiltonian formalism provides an alternative route to the same final equations of motion as we could find using either the Newtonian or Lagrangian approaches. (Taylor 2005, 532)

To emphasize, as said above, agreement on EOMs is stronger than agreement on observations, and physicists are aware of this fact: note how Taylor characterizes the equations of motion as clearly "one level above" observations:

> If we are very lucky, the equations of motion may have an analytic solution, but, even when they do not, they are the essential first step to understanding the solutions and they often suggest a starting point for an approximate solution. (Taylor 2005, 254)

I rest my case that students of physics are trained to treat the different formulations of classical mechanics not as mathematically distinct geometrical theories, but as "recipes" for achieving one ultimate goal: deriving the equations of motion, and that the latter are the "minimally theoretical" result that all dynamically equivalent theories converge on. In the following, I will argue that construing classical theories as fixed geometrical structures has been extremely misleading in recent debates on theoretical equivalence.

2.6 Conclusion

The two-tier character of physical theories (namely that they must have fixed state assignments and open-ended transition rules) has profound consequences for theoretical equivalence. As far as experimentally contentful theories are concerned, no one-to-one correspondence of models can be expected, for each model corresponds to an open-ended choice of

LEMP. In discussions of classical mechanics, in particular, traditional accounts have gone wrong both because they ignore the fixity of state assignments (and thus compare productive and parasitic theories to each other) and because they ignore the open-endedness of transition rules (and thus require too much of equivalent theories in a way that is inconsistent with the practitioner's proclamations).

3.0 Chapter 3: Essential Content in a Prescriptive-Dynamical Framework

3.1 Introduction

My arguments in the first two chapters relied quite centrally on the notion of "essential content". In this chapter I will delve more deeply into this notion, review the literature, and flesh out the relationship between the essential content of a theory (in the prescriptivist framework) and external reality in more detail.

In the following, Section 3.2 elaborates on the idea that state assignments are the essential / working parts of dynamical theories. First, I will use the case of Bohr's atomic theory in §3.2.1 to demonstrate my first two goals, namely that one has reason to believe that Bohr's theory has idle posits and that its state assignment rules are its essential content, regardless of debates surrounding realism. §3.2.2 is then devoted to generalizing beyond case studies my argument that state assignments are the generators of predictive success in dynamical theories. Next, in Section 3.3, I will argue my third point by offering an account of how state assignments are related to unobservable reality. This section is also divided into a specific and a general section: first, in §3.3.1, I show specifically that Bohr's states *track* the true dynamical states of the system in a precisely defined sense, but they need not in any sense *represent* or *refer to* anything in the system to be successful. Then in §3.3.2 I outline a generalized account of tracking and its most important features, including its relationship to predictive and referential success. Finally, in service of my fourth goal, Section 3.4 is devoted to a survey of existing accounts of selective realism. I submit that the necessity of referential transparency has been presumed (sometimes implicitly) by both proponents (§3.4.1) and critics (§3.4.2) of selective confirmation, and that this has caused the former group to struggle to produce a coherent account of essential content, and the latter to overstate the reach of their criticisms.

3.2 State Assignments Are the Essential Content

Not every component of a given empirical theory necessarily contributes to its empirical success. It is possible that the theory contains superfluous baggage that comes along for the ride but is not among the subset of posits that actually generate the predictive power of the theory. This extra baggage might be embellishment or scaffolding, a heuristic or "visualization" technique, put in place as a result of metaphysical prejudice, intellectual inertia, or whatever other personal or societal influence. To believe that one can differentiate between the "success-generators" of a theory and its superfluous baggage, and to have distinct epistemic attitudes towards each part, is to be a *selectivist*. The selectivist thus distinguishes between the "working" or "essential" posits of a theory and the "idle" or "merely presuppositional" ones, and recommends a favorable epistemic attitude only towards the former group of posits. That is why selectivism is sometimes referred to as the *divide et impera* move.[1]

Why be a selectivist? Selectivism is typically discussed as a savior of realism in the face of historical challenges such as the pessimistic induction (Laudan 1981) and the problem of unconceived alternatives (Stanford 2006). According to selectivists, the impressive novel success of a mature scientific theory should not and could not warrant a realist attitude towards *all* of the theory's posits, but only towards the subset of posits that actually contributed to bringing about the theory's success. This would block historical challenges *if it could be shown that these same parts also feature in subsequent theorizing*, for then the realist's success-to-truth inference would remain intact in the face of revolutionary theory change. This approach has been championed in particular by Kitcher (1993) and Psillos (1999), and subsequently many others (Leplin 1997; Niiniluoto 1999; Sankey 2001; Thagard

[1]To be clear, throughout this chapter as well as in the vast majority of the literature on selectivism, "success" always refers to *impressive novel predictions* (see Vickers 2013, 195 for a discussion of "unimpressive novel success"). Admittedly some authors (e.g. Chang 2003; Badino 2016) seem to work with other notions such as *explanatory* success, but that is more of an exception that proves the rule. For one to say that the prediction of a phenomenon P was a "novel" prediction by theory T, at the very least it would have to be the case that T was constructed without taking into account any data relevant to P or data based on which P could reasonably be expected. This might be either because the data had not been gathered yet (temporal novelty), or if it was, because the data was not deliberately baked into the theory (use novelty). It has been argued that novel predictive success is necessary for no-miracles-type arguments to go through (Saatsi 2009, 356, 360; cf. Vickers 2017a, 50 ff.). See Alai 2014 for a thorough discussion of the notion of novelty.

2000; 2007). Moreover, structural realism (as proposed by Worrall 1989 and others) can also be considered a form of selective realism.

However, although selectivism is almost always formulated as a defense of realism in the face of historical challenges, my first goal in this chapter is to argue that selectivism is independently motivated. In a way, my arguments in Ch. 1 already serve as independent grounds for believing that physical theories are often embellished or scaffolded by descriptive claims that are unnecessary for, if not detrimental to, the predictive success of the theory. There I argued that predictive success stems from the *prescriptive-dynamical* content of the theory. Since my arguments in Ch. 1 did not draw on realism or historical pessimism at all, they provide reasons to be a selectivist regardless of realist concerns. In the following, I will flesh this out both in general terms and through the specific example of Bohr's atomic model.

3.2.1 State assignments: the case of the Bohr model

3.2.1.1 Why Bohr's state assignments do all the work

Bohr (1913) posited that an atom is analogous to a mini-solar system, the nucleus being the "sun" and the electrons orbiting it being like "planets", but with one important restriction: of all the orbits that are classically permitted by Coulomb's inverse square law, the electron may occupy only those that satisfy the constraint

$$E_n = \frac{K}{n^2}, \tag{7}$$

where E is the binding energy of the electron, K is the Rydberg constant, and n is an integer known as the *principal quantum number*, which labels the discrete energy levels. Bohr further posited that the electron does not emit radiation unless it "jumps" from one orbit to another, in which case the difference of energy will be released as radiation of frequency

$$\omega_{nm} = \frac{1}{\hbar}(E_n - E_m). \tag{8}$$

Bohr was able to derive the emitted frequencies in the spectral lines of the hydrogen atom from these posits, which is not all that impressive in itself given that he constructed his theory

to account for the hydrogen spectrum. However, Bohr was also able to make impressive, novel predictions about the spectral lines of once-ionized helium, twice-ionized lithium, and other single-electron ions. These successes reportedly prompted Einstein to say "the theory of Bohr must then be right" (see Pais 1991, 154)! Yet of course Bohr's theory is certainly not anywhere nearly right about the mini-solar system and the classical orbits.

How did Bohr's theory achieve its success despite its radically mistaken world picture? Vickers (2012; 2013) lists the Bohr model as a potentially devastating problem for the selectivist. He remarks: "The crucial question for the [selective] realist is how it is possible for Bohr's theory to be so successful when we now know that it is so far off the mark (in the light of modern quantum theory)" (Vickers 2012, 10). Now, Vickers goes on to concede that a promising solution has already been proposed in the literature: Norton (2000) effectively argues that the essential or working posit of Bohr's model is the assumption of discrete energy states; the classical orbits play a purely heuristic role. Norton writes:

> The reduced form [of Bohr's model] eschews all talk of elliptical orbits other than in the domain of correspondence with classical theory... . No assumption is made or needed that these stationary states are elliptical orbits of some definite size and frequency of localised electrons. What is retained is that these states possess a definite energy. (Norton 2000, 86-87)

Norton's is essentially the same as the solution I will be recommending[2] (although Norton does not approach the problem from a prescriptive-dynamical standpoint).

One can convince oneself that the stationary states are all that Bohr needed in order to make his predictions regardless of realist anxieties. The experimental evidence to which Bohr was answering was that of spectral lines, which are characterized by two quantities: frequency (manifested in the relative separations of the lines and their colors) and intensity (manifested in the brightness of the lines). In order to predict the frequencies, all Bohr had to do was identify each spectral line as a *transition* and recommend a procedure for tagging each line with two integers n and m corresponding to the initial and final states of the transition. This procedure is encoded in Bohr's energy quantization conditions. The frequency is then given by the difference of energy between the two assigned states. In fact, it was known before Bohr that one can construct a series of wavelengths known as the "term

[2]See also Ghins 2014.

series" such that the wavelength of every spectral line can be obtained from the difference of two terms in the series. Bohr's innovation was to associate the terms with energy levels.

What about the intensities? They certainly could not be derived from state assignments alone. And in fact Bohr's theory does *not* provide any predictions about the intensities of the spectral lines. The intensities depend on the *transition amplitudes*, i.e. the probabilities associated with different transitions. The more likely a given transition is, the more frequently it will occur, thus increasing the brightness of the corresponding spectral line. As I've argued elsewhere (Kaveh 2014), the task of determining transition amplitudes and hence the intensities was delegated to the infamous Correspondence Principle. However, the Correspondence Principle was not a definite formula; it was an open-ended recipe for linking transition amplitudes to the Fourier coefficients of the orbits, and its exact form was to be determined experimentally. But as I have been arguing, this is precisely what one would expect from a dynamical theory! As I argued in Chapter 1, a physical theory consists of fixed prescriptions for state assignment and open-ended transition rules, with the latter being determined locally and empirically. There has been much puzzlement about the open-ended character of the Correspondence Principle, but if I'm right, that is simply how every physical theory operates. Sommerfeld's theory of the fine structure, proposed soon after, is another clear example in which the state assignments are the source of the theory's success (cf. Vickers 2012). Sommerfeld's theory had several "versions", each with a different set of transition rules, adapted to the experimental results obtained by Paschen around the same time (Kragh 1985, 71 ff.). Once again, in accordance with the prescriptive-dynamical account, we have a theory with fixed state assignments which function as the essential content and open-ended transition rules which are fitted to local empirical data under the presupposition of the state assignments.

The idea of keeping Bohr's state assignment rules and discarding the rest is not anachronistic and, contrary to Stanford's (2006) worries, was not "unintelligible" to scientists at the time. In the early 1920s, the idea of dispensing with classical orbits was gaining popularity among physicists including Bohr himself. Van Vleck discussed whether this "bold proposal" would "invalidate" the successes of Bohr's theory and concluded in the negative:

> Such successful applications, however, need not be forfeited if only we assume that the

> Bohr frequency condition and standard quantum conditions [i.e. the state assignment rules] retain their validity, even though the motions quantized by the latter are not in accord with ordinary dynamics... .
> These requirements are much less stringent than the restriction that the quantized orbits shall conform closely to the classical mechanics, and are, in fact, so very general that they do not determine at all definitely the form of the orbits. (Van Vleck 1926, 108-109)

To sum up, the empirical evidence was characterized by the two parameters of frequency and intensity. The former can be predicted through Bohr's state assignments, while the latter is local and empirical and *should not* be fixed (predicted) by the theory in the first place. Since there is no further empirical content, his state assignment rules exhaust "what was right" in Bohr's theory. The descriptive ontology of his theory (the entities, properties, and structures), on the other hand, was wrong through and through.

3.2.1.2 Vickers's objections

Vickers briefly raises two potential problems with the idea that Bohr's state assignments were his only essential posits. First of all, Vickers argues that Bohr's theory would not get off the ground if it were not for the way he sets his parameters by comparing them to the classical case in the classical limit:

> A possible objection to this story would be that Bohr still ... makes use of the assumption of elliptical orbits for very large quantum numbers... . Some might be tempted to claim that this is crucial to Bohr's derivation, but not even approximately true according to modern QM. (Vickers 2012, 10).

Vickers is right in that in his original presentation, Bohr *postulated* that in the limit of $n, m \gg n - m$, i.e. for orbits that have very large quantum numbers but are in the vicinity of each other (which is how one must characterize a classical Larmor radiation in Bohr's system), one must have

$$\omega_{nm} \approx \tau\omega \tag{9}$$

That is to say, Bohr posited that in the classical limit, the frequency of radiation must be an integer multiple of the mechanical frequency, as required by classical physics (Bohr 1913, 13; 1920, 429-430).

However, this does not support Vickers' objection for two reasons. First of all, contrary to what Vickers says, Bohr's postulate that radiation and orbital frequencies coincide in the classical limit does not imply that there are elliptical orbits in the classical limit. The assumption is explicitly only about frequencies, not orbits. Just as one can separate assumptions about the shape of the lower orbits from assumptions about their energy, so too one can distinguish assumptions about the shape of higher orbits and their frequencies.

In any case, there is a second problem with Vickers' first objection: as it turns out, Bohr did not actually need the assumption about the classical limit, because as Van Vleck later proved, equation 9 can be *derived* through the action-angle formalism, the quantization condition, and a few other trivial assumptions (Van Vleck 1924, footnote 4). Thus, equation 9 is not part of the essential content of Bohr's model because it is not a necessary premise in Bohr's argument.

The second objection Vickers considers is that in response to a point made by Fowler, Bohr argued that the nucleus is not exactly stationary, but "wobbles" due to the influence of the orbiting electron. Vickers argues that this is problematic for the selectivist because it crucially hinges on the assumption of real electron orbits. I do not see how this is the case, however. It seems that one could recover Bohr's results about the "wobbling" of the nucleus from the assumption that the electron has a non-trajectory-based presence around the nucleus but one whose net statistical effect on the nucleus is similar to that of an orbiting electron. Indeed, this is how the same effect is characterized in modern quantum mechanics.

More recently, Vickers (2018) has proposed a structural realist interpretation of the essential content of Bohr (and Sommerfeld). Vickers argues that there is structural continuity between the Bohr / Sommerfeld theories on the one hand and modern quantum mechanics on the other, on account of the fact that their equations look similar. The Schrödinger equation, written as

$$\frac{1}{2m}(\nabla S)^2 + \frac{\partial S}{\partial t} + U = \frac{i\hbar}{2m}\nabla^2 S \tag{10}$$

closely resembles the Hamilton-Jacobi equation and reduces to the latter in the limit $\hbar \to 0$. Moreover, the Bohr-Sommerfeld quantization conditions

$$\oint p dq = nh \tag{11}$$

differ from the modern semi-classical quantization conditions merely by a "Maslov term" $\frac{1}{2}h$ on the right hand side, which as it turns out plays no role in determining the energy states for systems of interest to Bohr and Sommerfeld (Vickers 2018, 15 ff.). Vickers thus draws on a "structural similarity in the limit"[3] account to explain the successes of Bohr and Sommerfeld's theories.

While I agree with the spirit of Vickers's solution, I find the talk of "structural similarity" to be unhelpful. As I will argue in §3.4.1.3, one of the main problems with structural realism is the fact that the notion of "structure" is too malleable to be contentful. If all one needs for structural continuity is that after sufficient symbolic manipulation two equations look partially similar with one having "excess structure" over the other, then structural continuity would appear too easily achievable. But more on this below. For now, let me simply point out that there is a less esoteric and less ambiguous reading of the equations above. The Schrödinger equation, the Hamilton-Jacobi equation, and the Bohr-Sommerfeld quantization conditions are examples of differential and integral equations. As I have been arguing throughout this dissertation, such equations are best interpreted as *state assignment rules*. This is of course most obvious in the case of the quantization conditions, which provide a prescription for assigning discrete states to a system. But any equation that provides final conditions for given initial and boundary conditions can be read as assigning dynamical states, and all differential equations function that way. Thus, I suggest that Vickers's "structural similarity" is better understood as one set of state assignments *tracking* another within certain constraints and resolution limits. More on this in §3.3.1.

I have argued that the state assignments were the only posits in the Bohr model that essentially generated its predictive success, i.e. its essential content. Note that my identification of essential content did not draw on any historical considerations, pessimistic or otherwise, and did not rely on any comparison between Bohr's model and subsequent theories. Selectivism is thus a well-motivated thesis independently of the realist's historical anxieties.

I do not believe the Bohr case is special. In the next subsection, I will present *general* arguments for the claim that state assignments are the success-generators of dynamical

[3]See Votsis 2011b, 4.

theories.

3.2.2 State assignments: the general account

In this section, I will generalize my claim that state assignment rules are the essential content of dynamical theories beyond the Bohr case.

First of all, Bohr's focus on state assignments as the essential content was carried over through Heisenberg's work to modern quantum mechanics, so much so that modern quantum mechanics may be considered the quintessential "ontology-less" dynamical theory. Healey (2017) has extensively argued that non-relativistic quantum mechanics should not be taken as describing the world, but rather as prescribing *credences* through its Hilbert states.[4] However, the identification of state assignment rules as essential content goes beyond quantum mechanical theories as well. I believe it can be generalized to all dynamical theories including classical mechanics.

Consider how a typical prediction is made in Newtonian mechanics. As explained in Chapter 1, it goes something like the following. One begins with a master formula such as $F = ma$, adds a suitable local empirical mediating principle (LEMP) such as $F = -kx$, and then puts the two together to obtain a differential equation such as $\ddot{x} = -\frac{k}{m}x$, called the *equation of motion*. The equation of motion is an input-output machine that returns final states of the form $(x_f, \dot{x}_f)$ given initial states of the same form: $(x_i, \dot{x}_i)$. But actually, one need not label the input "initial" and the output "final". The equation of motion returns values of $(x(t), \dot{x}(t))$ for all t given $(x(t_0), \dot{x}(t_0))$ for any t_0. In the case where the mediating principle is $F = -kx$, for example, letting $x(t_0) = x_0$ and $\dot{x}(t_0) = \dot{x}_0$ yields the following functional dependency of x on t:

$$x(t) = x_0 \cos\left(\sqrt{\frac{k}{m}}t\right) + \frac{\dot{x}_0}{\sqrt{\frac{k}{m}}} \sin\left(\sqrt{\frac{k}{m}}t\right). \tag{12}$$

One can use this equation straightforwardly to predict values of x for any given time t. The exact same steps must be followed when making predictions in cosmology, particle

[4]Healey's account therefore appears as a special case of the prescriptive-dynamical view, although Healey works with a slightly different (more doxastic) notion of prescription.

physics, solid state physics, thermodynamics, or any other modern physical theory (a table of examples was provided in Ch. 1). The following summarizes these steps:

Steps to making predictions in modern physics:

1. Write down a master formula (e.g. $F = ma$).
2. Add a suitable mediating principle (e.g. $F = -kx$).
3. Obtain an equation of motion (e.g. $\ddot{x} = -\frac{k}{m}x$).
4. Plug in initial and boundary conditions (e.g. $x(t_0) = x_0$, $\dot{x}(t_0) = \dot{x}_0$).
5. Obtain a functional dependency (e.g. $x(t) = x_0 \cos\left(\sqrt{\frac{k}{m}}t\right) + \frac{\dot{x}_0}{\sqrt{\frac{k}{m}}} \sin\left(\sqrt{\frac{k}{m}}t\right)$).
6. Plug in specific values of the independent variable (e.g. t) to make predictions about the dependent variable (e.g. $x(t)$).

What is the essential content in such a derivation? What assumption(s) "fuel(s)" this prediction and what is their essential content? The best way to answer this question is to start at the bottom, i.e. at the step closest to observation, and move up the chain of inference to see what is the minimum set of assumptions we would have to keep to preserve the prediction.[5] At the closest level to observation, there is the functional dependency (e.g. $x(t)$). This is of course what fuels the predictions *for the system to which it applies*, but it only applies to very specific systems that exemplify the specified initial conditions. It would not make correct predictions for any other systems. As such, the functional dependency is not a *theory*. In fact, if one were merely interested in a single system, the functional dependency could be derived by fitting a trendline to the data; there would be no need for a theory. The essential content must therefore be not the functional dependency itself, but whatever fuels the derivation of it.

One level above, we have the equation of motion plus the initial conditions. This is what entails the functional dependency, but it cannot be the essential content of the theory either, because insofar as it is characterized by particular initial and boundary conditions, it is just as specific as the functional dependency itself. Besides, even the equation of motion itself is too specific. While it would be valid for a wide range of initial and boundary conditions, it

[5]This is in essence the method recommended by Cordero (2015, 11) for finding the essential content.

can only be a model for systems that are characterized by a particular mediating principle (e.g. $F = -kx$ for harmonic oscillators). By contrast, what we are looking for is the essential content of *the general theory*, not that of a specific model of a specific (type of) system.

Moving up one more level, we arrive at the conjunction of the master formula and the mediating principle. Now, this conjunction has no more generality than the equation of motion, seeing as the cause of specificity of the latter is the mediating principle itself. However, moving to this level opens the door for generalizing: rather than having a single mediating principle, one can take the essential content to be the conjunction of the master formula on the one hand and a massive disjunction of mediating principles, each suitable for a particular type of system, on the other. For each given system, one chooses the appropriate disjunct, and things roll from there. It might seem that this is the end of the line; there is no set of assumptions that can yield the same prediction but is weaker than the conjunction of the master formula with the massive disjunction of mediating principles. Removing either premise would render the inference invalid.

However, if my claims in Ch. 1 are true, this is not correct. *There is no massive disjunction that provides an exhaustive list of all physically relevant mediating principles.* Rather, the mediating principle is open-ended and must be "reverse engineered" from local empirical data *under the presupposition of the master formula.* Therefore, in my account the steps above should be revised as follows.

Steps to making predictions in modern physics (revised):

1. Write down a master formula (e.g. $F = ma$).
2. Fit the master formula to local empirical data to obtain a LEMP (e.g. $F = -kx$).
3. Conjoin the resulting mediating principle with the master formula to obtain an equation of motion (e.g. $\ddot{x} = -\frac{k}{m}x$)
4. ...

Since the mediating principle is derivative from the master formula and empirical data, it is not an *essential* posit of its own. The only essential posit appears to be the master formula itself.[6]

[6]Note that none of my arguments so far in this section draw on the dynamical systems framework

The only question now is: what is the content of the master formula? Consider $F = ma$ again. One way (perhaps the most straightforward way and certainly the most common way) to read the content of this formula is as follows. $F = ma$ tells us that there is an entity called force whose magnitude when acting upon a system stands in such and such a relationship to two of the system's properties, namely mass and acceleration. In this way, the essential content of Newtonian mechanics would contain assertions about the properties and relations of various entities such as forces, masses, and so on.

But this cannot be right for a number of reasons. First of all, as I argued in Ch. 1, $F = ma$ (or any other master formula) is quite contentless when taken as a descriptive statement: as long as the system can be assigned mass and acceleration, surely there will be *some* function whose value at each moment equals the product of mass and acceleration. That statement by itself certainly cannot generate any predictions. Besides, the essential content of Newtonian mechanics better not involve entities such as forces, for these entities are absent from Lagrangian and Hamiltonian mechanics, which make the same predictions as Newton's laws about most classical systems. It would therefore be much more reassuring for the selectivist if the essential content of these three theories was the same. But if they are to be the same, they cannot include forces.

I provided extensive arguments in Ch. 1 and numerous quotes from physics textbooks in Ch. 2 to the effect that $F = ma$ (and other master formulae) must be read as *prescriptions*. This saves the formula from the charge of contentlessness. $F = ma$ is essentially an injunction that tasks us to form differential equations of the form $\ddot{x} = f(x, \dot{x}, ...)$ and search for a suitable f in each local context. The left-hand side of this differential equation fixes what our initial / boundary conditions, i.e. our *state assignments*, must be. In this case, they are $(x, \dot{x})$, due to the fact that the equation is second-order in x. The right-hand side, on the other hand, is given by a local empirical mediating principle (LEMP) and encodes the transition rules for states of this form. As said, this part is open-ended and must be reverse engineered from empirical data by presupposing the prescribed state assignments. So really the only fixed and essential content of $F = ma$ is that it commands us to assign states of the form $(x, \dot{x})$. Thus, my general claim is the following:

(although they do draw on prescriptivism).

Claim: The essential content of a dynamical theory is its master forumla, whose essential (prescriptive) content is a set of state assignment rules.

So far I have based my arguments on how predictions are actually made in the practice of physics. One can arrive at the same conclusion (that state assignments are the success-generators) by considering in most abstract terms what one would need in order to make predictions. Hempel was one of the earliest to study the nature of predictions and the role of theory in them. Hempel (1958) claimed that the role of a theory is to provide a "secure link" between "observational antecedents" and "observational consequents". I agree with Hempel that predictions paradigmatically consist of going from conditions in one time and place to conditions in other times and places.[7] However, Hempel's statement regarding the role of theory in predictions is wrong in a subtle but crucial way. Textual evidence strongly suggests that for Hempel "observational antecedents" and "observational consequents" mean initial and final conditions, respectively, and the "secure link" refers to transition rules. But if that is the case, then Hempel has got it completely backwards: the role of the theory is *not* to provide transition rules between observational antecedents and observational consequents, but rather to tell us *what the observational antecedents must be to begin with.* Once this is provided, the transition rules are in fact not theoretical at all but rather, as I have been arguing, derived from local and empirical data. Indeed directly reading the transition rules off empirical data is the practice Hempel recommends as a replacement for the theoretician's method, namely that of going through theory to make predictions. This is where the theoretician's "dilemma" stems from: why keep the theory around, if you can simply record the transitions between observational antecedents and consequents directly? But as said, what Hempel calls the theoretician's method is fictional;

[7]To be sure, not every prediction directly involves anticipation of a final condition or specification of initial and boundary conditions. Take, for instance, the "prediction" that every planet sweeps equal areas in equal amounts of time, which may be derived from Newton's inverse-square formula. On the face of it, this is not a prediction of a final state given some initial state, but rather a formula relating swept areas to times. However, it is clear that the driver of this prediction is Newton's differential equation $\ddot{x} = \frac{F(x)}{m}$, which spits out the position and velocity of a planet at a desired time given its position and velocity at some other time. It is this dynamics that gives rise to the area law, for it is the dynamics that determines how far and how fast a planet will travel given various starting points, and it is the regularities in these inputs and outputs that is distilled into the area law. As such, even if the prediction does not appear dynamical on the surface, it nonetheless originates in a dynamical rule. Therefore, one would expect a study of input-output dynamics to be most urgent and most informative in a discussion about predictions.

no physicist uses the theory to derive transition rules, for those *are* recorded empirically. But to derive the transition rules from empirical data, one must know what to keep track of as the "antecedents" and "consequents", and for this one needs the theory.

Suppose one is interfacing with a classical system. At each moment the system is characterized by an indefinite number of measurable quantities: x, $\dot{x}$, $\ddot{x}$, $(x+\dot{x})\ddot{x}$, $\frac{x}{\dot{x}}$, and so on. Which (collection) of these quantities should one keep track of as the appropriate "observational antecedent"? This is not trivial at all. Many centuries of rigorous investigation were needed before Newton essentially realized that we must keep track of $(x, \dot{x})$. Once this was discovered, the transition rules were simply a matter of paying careful attention to data (see my remarks on Newton's study of gravity in Ch. 1). In short, Hempel seems to assume that the specification of observational antecedents is trivial, but it is in fact *the most important* contribution of a theory.

Once again we arrive at the same conclusion: the state assignment rules are the sole source of the miracles that dynamical theories pull off. In the following, I will examine the question of what these posits can be said to "latch on to".

3.3 Tracking: Realism without Reference

In this section, I will concern myself with how theoretical state assignments of a successful dynamical theory relate to the true states of the system. As said above, I shall call this relation *tracking* and contrast it with its dominant alternative, *referential success*. Given that tracking is a systematic relation between the unobservables of the theory and elements of reality, and one that explains the predictive success of the theory in a non-miraculous way, this view answers to the No Miracles argument and may thus be considered a form of "realism". On the other hand, if by "realism" one intends any view in which the theoretical unobservables are "real", i.e. transparently referential, then my view is a form of antirealism. Labels aside, I claim that *privileged epistemic and metaphysical status can be conferred upon certain unobservable terms without appealing to referential success or any other referentially transparent notion.* As before, I will proceed with a specific case study before presenting a

general account.

3.3.1 Tracking: the case of the Bohr model

I shall argue that the Bohr case vividly demonstrates the bankruptcy of referentially transparent notions such as "referential success", while still allowing for a systematic (albeit referentially opaque) theory-world relation to explain the theory's predictive success: Bohr's states *track* the true states of the system, but need not *represent* or *refer to* them. To this end, I will draw on a comparison of two experimental cases, namely the case of the hydrogen atom (or any single-electron ion such as He^{+}, Li^{++}, etc.) and that of orthohelium, respectively. Importantly, the former was considered a *confirmation* of Bohr's theory while the latter was taken as a *disconfirmation*. As such, studying the similarities and differences between the two cases is the best way to get a handle on what makes a false theory predictively (un)successful.

Before examining the cases, allow me to say a few words about the methodology I will employ below. My approach is to examine the relationship between the Bohr energy states and those of modern quantum mechanics, governed by the Schrödinger equation. One might wonder why we should care about this comparison, given that ultimately we are interested in how Bohr's theory relates to *the truth*, not to any other theory. To be sure, the Schrödinger equation does not provide us with the "true states" of the system, seeing as it is ultimately a false theory itself. Nevertheless, if it could be shown that referential success is the wrong way to characterize the relation between Bohr states and Schrödinger states, then it would be extremely difficult to argue that the Bohr states refer to anything in reality, unless one believes that modern quantum mechanics is somehow farther from the truth than old quantum theory. That is my negative claim. My positive claim is that this comparison will give us insight into what relationship between the terms of the Bohr theory and those of Schrödinger allows the former to be successful, despite how radically mistaken it is by the lights of the latter. If whatever relationship turns out to do the work is generalizable to other cases, then one would expect the Schrödinger theory to stand in a similar relationship to quantum field theory, and the latter to whatever future theory replaces it, and so on. And

if this relationship is a transitive one, then given the right context, Bohr's theory stands in the same relationship to quantum field theory, etc., and in the end also to "the ultimate true theory" (if I may be allowed to use that phrase) which by assumption perfectly mirrors reality. Thus, assuming a generalizable and transitive relation exists, examining the relation between Bohr and Schrödinger models would be a promising way to get insight into the relation between the Bohr model and the truth. As said, I will argue that the relation in question is tracking (defined below), and it is indeed both transitive and generalizable beyond this case.

3.3.1.1 The two cases introduced

The two cases are shown below. Figure 1 illustrates the relationship between Bohr and Schrödinger states for a hydrogen atom (a similar pattern would emerge for any single-electron ion).[8] In such contexts, $E_n = \frac{K}{n^2}$ is a decent approximation to the energy levels, and therefore Bohr states line up pretty nicely with Schrödinger states. I say "pretty nicely" because there is a *fine-structure* which, from the contemporary point of view, is due to spin-orbit interactions and relativistic corrections. This results in some separation between states of the same principal quantum number n but different azimuthal quantum numbers. And since these are not accounted for in Bohr's theory, this causes misalignment between the two sets of states. Note that while the origin of fine structure was not known at the time, its existence had been experimentally established since the 1890s, and the inability of Bohr's original model to account for them was not in dispute when Bohr proposed his theory (see Kragh 1985).

Next consider a similar comparison between Bohr and Schrödinger states for an ortho-helium atom, a neutral helium atom ($Z = 2$) in the spin-1 mode. This is shown in Figure 2 (note that fine structure has been suppressed in this figure).[9] Orthohelium is to be contrasted with parahelium, a neutral helium atom in the spin-0 mode. The main difference between the two is that unlike parahelium, the second electron in orthohelium is exiled to $n \geq 2$ and cannot decay to the ground state. The figure for parahelium would therefore look

[8] For precise values of the energy layers see Nanni 2015, 60.

[9] Figure based on Rohlf 1994, 267, Figure 9-11.

quite similar, except $2s$ would be pushed 0.8 eV higher, and more importantly there would be a $1s$ state, because the second electron in parahelium is allowed to decay into the ground state where the first electron is. From the contemporary point of view, the latter fact is due to the relative orientation of the electrons' spins. The two electrons in an orthohelium atom have parallel spins and are therefore excluded from the same quantum state due to fermionic anti-symmetry (the "exclusion principle"). In parahelium, by contrast, the two electrons have anti-parallel spin and thus prefer to be in the ground state together. One says that orthohelium's $2p$ is *meta-stable*, which is to say that an orthohelium electron in $2p$ does not spontaneously decay to the $1s$ ground state. In pure samples of orthohelium, the $2p \rightarrow 1s$ transition is not found among the spectral lines. I shall differentiate an orthohelium state from the analogous parahelium state by writing subscripts o for orthohelium and p for parahelium, e.g. $2s_o$ vs. $2s_p$, $2p_o$ vs. $2p_p$, etc.[10]

3.3.1.2 In search of a difference-maker between the two cases

The crucial question is this: How is the case of neutral orthohelium different from the case of hydrogen or ionized helium? What difference could explain the fact that one is widely considered a confirmation of Bohr's model but the other a disconfirmation? In what sense were Bohr's state assignments to single-electron atoms and ions "onto something" but not his state assignments to orthohelium? What did they "latch on to" in the former case that they failed to capture in the latter?

First of all, contemporary textbooks of physics and chemistry are of no help here. It is commonplace in these texts to claim that Bohr's theory works great for hydrogen and single-electron (or "hydrogen-like") ions but fails miserably as soon as it is applied to helium or any other system with more than one electron. What is *not* easy to find in these texts is a good reason for these assertions. Sometimes it is said that Bohr's theory fails in the case of helium because it cannot take electron-electron interactions into account (OpenStax 2016, §6.2). This is clearly false: in fact, Bohr discusses such interactions in the second part

[10]The contemporary notation is 2^3s for orthohelium and 2^1s for parahelium, because the former turns out to have a triplet fine structure while the latter is a singlet with no fine structure. But at the time, 2^3s was perceived as a *doublet* (Gearhart 2017), so the contemporary notation would be too anachronistic here.

of his famous trilogy (1913), in which he presented his model for the very first time. Or take Bohr's letter to *Nature* titled "Atomic Structure" in which Bohr explicitly considered electron-electron interactions, including the possibility of orbits penetrating below the lower shells (Bohr 1921; see also Darrigol, 152-153). And of course the many helium models around this time gave different results for ionization energy mainly because of differences in their electron-electron interactions. Other textbooks claim that Bohr's theory fails for multi-electron atoms and ions because in this theory the binding energy only depends on the principal quantum number n whereas experimentally it also depends on the azimuthal quantum number l which indicates angular momentum, and/or other quantum numbers such as spin that are missing from Bohr's theory (Uppal 2006, 426). But this cannot be the difference-maker either, for two reasons: first of all, one can add further quantum numbers to Bohr's n, and this is indeed what was done by Sommerfeld and others immediately after Bohr's proposal. Secondly, as we saw above, in the case of hydrogen and ionized helium, too, the energy levels depend on angular momentum due to fine structure. There is therefore no meaningful difference between the two cases in terms of lack of further quantum numbers. Once again, whatever the problem is, it must be one that is present in the orthohelium case but absent from hydrogen and ionized helium cases.

Let us see if we can find a meaningful difference between the two cases by examining the two figures. At first glance, the two figures look quite similar. Obviously the Bohr states do not perfectly match up with any of the Schrödinger states in either case. In both cases one sees more spectral lines than expected from Bohr's theory, and one sees them in wrong locations. For instance, a transition such as $3p_o \rightarrow 2s_o$ in orthohelium would release more energy, and thus a higher frequency, than $3s_o \rightarrow 2p_o$. Yet both of those transitions would map onto $3 \rightarrow 2$ for Bohr. This means, in terms of experimental data, that one sees more spectral lines with more (and different) colors than promised by Bohr's model. The same can be said about transitions among the fine structure states. Those are also discernible and as said above, were accepted at the time as being beyond the margin of error.

So, in both cases, Bohr's theory "lumps together" physically distinct states. In both cases it is *strictly false* but *approximately true*. To be sure, the approximations are quantitatively much better in the case of hydrogen. Indeed, the sub-layer separations in Figure 1 are orders

of magnitude smaller than the average of their energy levels, while the separations in Figure 2 are comparable to the average energy values. However, the difference between the two cases could not simply be a quantitative matter. First of all, I have only shown the states according to the original 1913 theory. If one were to add the azimuthal quantum number k to Bohr's theory (a la Sommerfeld), the quantitative agreement with data would significantly improve. Secondly, as I explain in the historical Appendix B, quantitative discrepancies in helium were considered resolvable at the time. Thirdly and most importantly, as long as both deviations are considered statistically significant, the mere quantitative difference cannot explain why one case counts as success and the other as failure. At the end of the day, if the separations cannot be attributed to experimental error, it does not matter how big the discrepancy is: in both cases Bohr's theory strictly contradicts experimental data in that it has the wrong number of terms and has them at the wrong places. As Kragh explains:

> The observed fine structure could have presented a real problem to Bohr's theory... . However, it was not regarded as an anomaly and not allowed to impede the rapid acceptance of Bohr's theory. (Kragh 1985, 69)

Yet as we shall see, experiments on helium in the 1910 and 20s were indeed the death knell of Bohr's theory. We are back at the question: why did physicists consider the first one a success and the second a failure?

3.3.1.3 Tracking: the difference-maker

As we saw, in both cases Bohr's theory "lumps together" what are in reality distinct states. It defines coarse-grain "equivalence classes" of underlying states, if you will. The question is what makes the coarse-graining acceptable in the case of hydrogen and ionized helium but unacceptable in the case of neutral helium. I claim that the two cases cannot be distinguished *unless one pays attention to the underlying transition rules.* In other words, my claim is that the equivalence classes cannot be justified merely in terms of the form of the vertical *mapping* between Schrödinger states and Bohr states, i.e. the accuracy with which Bohr states "line up against" or "approximate" or "represent" Schrödinger states. To determine which states can belong in the same equivalence class one must also study the

manner in which those states *evolve*. In short, my diagnosis is this: the fundamental reason for the failure of Bohr's model vis-a-vis helium is that *it lumps together states that can be prepared separately and shown to transition in a distinguishable manner according to the very same theory*. Allow me to explain.

When the spectral lines of sources believed to contain helium were first studied, it was realized that upon constructing the "term series" (what Bohr later interpreted as energy levels) for the spectral lines, one obtains two separate sets of terms (energy levels), as it were belonging to two different elements. That is, one finds no transitions "across" the two sets of states. The two elements were given different but similar names. One was referred to simply as "helium" (later "orthohelium") and the other as "parahelium" or "parhelium" (Mehra and Rechenberg 1982, 398). What we now know as 2^3s was at this time considered the *ground state* of (ortho)helium. But soon it was realized that the two "elements" are really two excited states of the same element, now called helium. The experimental findings were thus reinterpreted as indicating a "no combination rule": orthohelium and parahelium states never transition among themselves. For example, while both $2p_o \to 2s_o$ and $2p_p \to 2s_p$ happen quite frequently, there are never cross-transitions of form $2p_p \to 2s_o$ and the like. It was accepted that any adequate model of helium must entail the no combination rule.[11]

I suggest that the problem with Bohr's theory in the case of helium was precisely that its state assignments were incompatible with the no combination rule. Extensive historical arguments for this claim are presented in the Appendix B (see also Darrigol 1992, 177). For the purposes of this section, we need only the following. Born and Heisenberg showed in 1923 that Bohr's state assignment rules force us to put orthohelium states in the same equivalence class as those of parahelium. For instance, the $2p$ states of orthohelium and parahelium are lumped together in the same set of stationary states. Since the two states end up having the same quantum numbers and both belong to the same set of stationary states, Bohr's theory does not have the tools to allow para- and ortho-states to transition differently. But as we saw, they in fact do. In the simplest case, $2p_o$ always transitions to $2s_o$ while $2p_p$ transitions to $2s_p$ and frequently to $1s$. Therefore, two states that Bohr's theory lumps together in the

[11]Recent experimental findings indicate some rare transitions across para- and ortho-states (see NIST Helium Tables), but the fact that these have such extremely low amplitudes still needs to be accommodated in the theory.

same equivalence class, in this case $2p_o$ and $2p_p$, transition to two other states, i.e. $2s_o$ and $1s$, respectively, *that the same theory recognizes as distinct*.

Note the difference with the case of hydrogen. The various hydrogen states that Bohr's theory lumps together are relatively well-behaved: for the most part, states that Bohr's theory lumps together transition to other states that the theory also recognizes as the same. The only exception to this is the $3 \rightarrow 1$ transition, which is only allowed for $3p \rightarrow 1s$ (suppressing fine structure) but not for the other $n = 3$ states. But as I shall explain in §3.3.2.2, this would only be a problem for Bohr's theory if i) one were not allowed to add the azimuthal quantum number, and ii) one were able to prepare $3p$ separately from the other $n = 3$ states. Since neither condition was true at the time, the coarse-graining of various hydrogen states mostly respects the way the underlying states transition.

The failure of Bohr's theory for helium, I claim, is a special case of why any dynamical theory might fail to predict: i.e. that there are several states that the theory recognizes as being in the same equivalence class, but when one prepares them separately, each regularly transitions to final states that the theory itself recognizes as distinct. This shows that the initial states should not have been put in the same equivalence class, and therefore that the theory's state assignments are faulty. By contrast, when states within the same theoretical equivalence class cannot be shown to transition distinguishably, I will say that the theory *tracks* the true states of the system within the constraints and resolution limits of the context.

> *Tracking:* A set of states S *tracks* another set of states S' iff there is a mapping from S' to S (inducing equivalence classes on S') such that members of each equivalence class defined on S' transition in a manner that is indistinguishable according to the same state assignment rules *at least within a range of contexts*.

Bohr's state assignments thus do not fail in the case of helium because they fail to represent something in the system: in one straightforward sense of the term, in both cases they *do* represent groups of underlying states with the same n, and in another sense of the term, namely one-to-one correspondence, they do not represent anything in either case. Rather, the state assignments fail because they are maps of groups of underlying states that transition distinguishably according to the very same way of representing them. Put differently, the problem with Bohr's theory vis-a-vis helium cannot be cashed out in terms of

the properties of the "mapping" between Bohr states and true states; the problem rather lies in the way this mapping interacts with the inherent transition rules of the system. Focusing on the static referential / representational relation alone will not give us a rich enough account to explain predictive success or failure of dynamical theories.

I shall have more to say about tracking in §3.3.2. But first let me elaborate on the last claim I made, namely that referential success is not sufficient for explaining predictive success and failure.

3.3.1.4 Why referential success is not the difference-maker

Suppose there were something the Bohr states referred to / represented in the case of hydrogen, something that could be said to have generated the theory's success. What would this referent be? The most obvious candidate would be the principal quantum number n. Bohr's theory works as much as it does, the idea goes, because it successfully refers to the principal quantum number, which has been "retained" in subsequent theories as one member of a larger set of quantum numbers. But this victory would be short-lived for the proponent of referential success, because this move would make referential success incapable of explaining the failure of Bohr's theory. For if Bohr states refer to the principal quantum number in one case, they must do so *in both cases*, and of course in both cases n has been "retained" in the superseding theory. Thus, Bohr's theory would be referentially successful in both cases but predictively successful only in one, which means that referential success cannot explain predictive success.

On the flip side, both in the case of fine structure and in the case of meta-stable helium, the main cause of deviation from Bohr's theory (according to current theory) is the omission of spin. In neither case does Bohr's theory represent spin in any way, yet it works for hydrogen and single-electron ions despite its failure to refer. Once again, referential success / failure cannot differentiate the two cases.

The only option for the proponent of referential success is to say that, whatever Bohr's states refer to, they are referential in Figure 1 but not in Figure 2. One would have to say, for example, that Bohr's $n = 3$ state refers to $3s_{\frac{1}{2}}$, $3p_{\frac{1}{2}}$, $3p_{\frac{3}{2}}$, $3d_{\frac{3}{2}}$, and $3d_{\frac{5}{2}}$ energy states

in a helium ion, but that the same state suddenly refers to *nothing* if the ion acquires a second electron and becomes neutral. This strikes me as a bizarre manner of speaking, but regardless, the question would then be: what would account for the fact that Bohr's states fail to refer in the case of orthohelium? The proponent of referential success must provide an *independent* reason for thinking so, or else this solution is utterly ad hoc. One should be able to determine, without knowing which case was considered a predictive success and which a failure, what (entity, property, or structure) the Bohr states refer to in the case of hydrogen-like systems, and then show that the same entity, property, or structure does not exist in orthohelium, and hence that Bohr's terms are non-referential in that case. As I just argued, this referent most likely cannot be the principal quantum number, for then the theory would be referential in both cases. The proponent of referential success must come up with an alternative referent that exists in hydrogen but is absent from orthohelium. I do not know what that alternative could be, and until one such alternative has been justified, I see little hope for referential success as an explanation of predictive success.

But let us assume one were able to pull off the claim that Bohr's states refer to something in ionized helium but not in neutral helium. Far from helping the referentialist, this would ironically *weaken* the notion of referential success significantly. For it would then appear that whatever Bohr's theory is representing about hydrogen is a quite volatile and unstable entity (or property or structure or what have you), seeing as it immediately disappears upon the second electron's approach. Thus, I am inclined to conclude, whether the theory is successful is not determined by whether it represents anything in the system, at least not anything robust.

3.3.2 Tracking: the general account

In this section, I will discuss the generic properties of the tracking relation and its relationship to predictive as well as referential success.

3.3.2.1 Tracking defined

To generalize, my claim is that the predictive success of a dynamical theory depends on a kind of "cooperation" or "compatibility" between the coarse-graining procedure and the underlying dynamics of the system. More formally, predictive success hinges on a form of compatibility between the vertical mapping that projects the true states onto theoretical states on the one hand, and the horizontal transitions among the underlying states on the other. I have been calling this *tracking*. Let me reproduce the definition here.

> *Tracking:* A set of states S *tracks* another set of states S' iff there is a mapping from S' to S (inducing equivalence classes on S') such that members of each equivalence class defined on S' transition in a manner that is indistinguishable according to the same state assignment rules *at least within a range of contexts*.

The idea of tracking is not an obscure philosophical invention. It has appeared in the writings of at least one prominent physicist. Let me quote the opening paragraph of Hertz's *Principles of Mechanics*:

> In endeavouring thus to draw inferences as to the future from the past, we always adopt the following process. We form for ourselves images or symbols of external objects; and the form which we give them is such that *the necessary consequents of the images in thought are always the images of the necessary consequents in nature of the things pictured*... We are thus enabled to be in advance of the facts, and to decide as to present affairs in accordance with the insight so obtained. The images which we here speak of are our conceptions of things. With the things themselves they are in conformity in one important respect, namely, in satisfying the above-mentioned requirement. For our purposes *it is not necessary that they should be in conformity with the things in any other respect whatever*. As a matter of fact, we do not know, nor have we any means of knowing, whether our conceptions of things are in conformity with them in any other than this one fundamental respect. (Hertz 1894, 1 – emphasis added)

It is not difficult to see how this passage simultaneously describes the process of tracking and denounces any referential inferences that may be made from it. Taking Hertz's "imaging" to be our "mapping" between lower- and higher-level states, and his "necessary consequent" to be the final state to which a given state transitions, Hertz is basically demanding that if the initial states are "lumped together" by the theory, then the final states also be lumped together by the theory.

Actually, what Hertz describes is a necessary but not sufficient condition for tracking,

which I shall call *co-transitioning*. This requires only that states within the same equivalence class also transition to the same final equivalence class. Rosaler (2015a, 61; 2015b, 11) has recently given co-transitioning a formal definition: let m be the map that takes us from true states to theoretical states, D the *true* transition operator that takes us from true initial states to true final states, and D' the *theoretical* transition operator taking theoretical initial states to final ones. Then co-transitioning can be neatly written as: $m[D(S)] \approx D'[m(S)]$. While always necessary, co-transitioning is sufficient for tracking only if i) the underlying dynamics is deterministic and ii) the differences of magnitude among states of the same equivalence class cannot be resolved by our instruments. For otherwise two states that co-transition might nevertheless result in distinguishable processes if the two transitions have discernibly different amplitudes or magnitudes. But if (i) holds, then the only transition amplitudes are 0 and 1 and thus there cannot be a difference of amplitude without a failure of co-transitioning. And if (ii) holds, one need not worry about differences of magnitude.

A note about the generality of tracking: my focus has been the Bohr-Schrödinger case, but tracking can be applied much more broadly within dynamical systems. Some work in this area has already been done. A relation much like co-transitioning was introduced in Wallace (2012, 54), who calls it "instantiation", and recommends it as the way to think about the quantum-classical relationship (essentially showing that classical mechanics tracks quantum states in the appropriate context). The concept has since been popularized by Rosaler (2015a; 2015b), who refers to it as "DS reduction" (or occasionally "tracking"), and argues that it serves as a better account of theory reduction than the traditional alternatives. Yoshimi (2012) has recently argued that this "compatibility condition" is helpful in addressing the relation between brain-theories and higher-level psychology. See also Giunti's discussions of this form of reduction under "emulation" (2006).[12]

But the idea of tracking applies even more broadly. One can show, for example, that spin states track Dirac spinors under suitable constraints and resolution limits. I also believe that many of the most prominent historical examples in the literature can be analyzed in the manner above: one could interpret theories of phlogiston, caloric, and ether, for instance,

[12]While all of these authors have appealed to this restricted notion of tracking as a promising *inter-theory* relationship, I have argued that it can be put to use also as a theory-world relationship and a much better alternative to referential success.

as assigning states to chemical substances, thermodynamic systems, and light, respectively, and following their transitions under the guidance of these state assignments. One can then explain the success of these theories by saying that, within certain constraints and resolution limits, these state assignments were compatible with the underlying transitions of said systems, i.e. that they tracked them. A detailed treatment of these cases is beyond this chapter. So while more work needs to be done to explore further applications of tracking, I hope my arguments have at least shown the promise of the idea.

Hertz's passage implies that tracking is necessary and sufficient for predictive success. As we shall see in the next section, this is on the right track but not quite accurate (even if tracking were reduced to co-transitioning). In the following, I will examine the necessity and sufficiency of tracking for predictive success.

3.3.2.2 Tracking and predictive success

First of all, tracking is sufficient for the predictive success of a dynamical theory. To test a prediction, one prepares the system in a state recognized by the theory, allows the state to evolve, and then detects the final state also in accordance with what the theory assigns. Now, unless one has arrived at "the ultimate true theory", the procedure the scientist considers to be the preparation of "the same state" is bound to prepare several states that are in reality distinct but assigned the same state by the theory. If, however, a tracking relation obtains between the true and theoretical states, then the scientist will never have to know that they are preparing different states under the same label, because whenever they prepare "the same initial state", they will also get "the same final state". Equipped with these state assignments and the empirically recorded transition rules, one can then use the theory in the future to predict, reliably and without contradiction, what final product will result from a given preparation procedure. Tracking is therefore sufficient for predictive success.

But tracking is *not* always strictly necessary for success. In some cases failure of tracking amounts to direct disconfirmation, while in others it indicates *potential problems* and suggests the *possibility* of better alternatives for the theory. In the latter cases, if such better alternatives are found, then the theory will be superseded; otherwise, it will remain

in practice. What determines whether failure of tracking is truly devastating for predictive success or merely indicates potential problems? In most cases the answer is simple: *it depends on whether individual states in the same equivalence class can be separately prepared and observed*.

To see this, consider two underlying ("true") states S_1 and S_2 that fail to co-transition. That is, suppose both S_1 and S_2 are mapped to the same theoretical state S_T, and let the transition rules be $S_1 \to S_3$ and $S_2 \to S_4$, and suppose S_3 maps to some S'_T and S_4 to some S''_T such that $S'_T \neq S''_T$. Since the two states fail to co-transition, tracking fails. Is this a problem for the theory? First of all, it certainly will be a problem if S_1 and S_2 can transition between themselves, for instance if $S_3 = S_2$, for in that case there will be a transition that the theory simply cannot accommodate. Consider for example the bare 1913 theory of Bohr with just the principal quantum number. This theory would not be able to account for orthohelium's $2p \to 2s$ decay, because the two states would both be mapped to Bohr's $n = 2$, and there is no $2 \to 2$ in this theory (certainly not one that can release energy).

What if there are no such in-class transitions? In that case, the failure of tracking would only be problematic *if the experimenter is able to prepare S_1 and S_2 separately* through different procedures. This is because in that case one could easily demonstrate that what the theory designates as S_T is in fact not a unique state of the system, for it can be reliably shown to transition differently depending on how it is prepared: namely to S'_T in one case and S''_T in the other. Note that the experimenter does not need to *be aware* that two distinct states are being prepared. In fact, if the experimenter is an adherent of the theory in question, then they would probably *not* know this, as they are inclined to lump both states into one. The point is that if there are two prescriptions for preparing S_T but each way *actually* results in a different initial state, and if the two states thus prepared transition distinguishably, then the theory is clearly coarse-graining too much.

For instance, given the relatively large difference of energy between them, it may be possible to prepare an orthohelium atom one time in $4s$ and another time in $4f$ in two separate experiments by applying different amounts of excitation energy in the discharge tube. Even if the experimenter does not already know this, the data will prove that two distinct states are being prepared in the two experiments because one would transition to

$2p$ while the other would show no spectral line indicating a $4 \rightarrow 2$ transition (NIST Helium Tables).

But now let us assume the experimenter is *not* able to prepare S_1 and S_2 separately. That is, suppose every preparation procedure at our disposal will be equally likely to bring the system to either S_1 or S_2. In that case, and as long as there are no in-class transitions ($S_1 \not\leftrightarrow S_2$), the data will show that S_T has *indeterministic transition rules*: sometimes S_T transitions to S'_T and sometimes to S''_T. The experimenter will be able to record the individual amplitudes for each transition and construct a probability matrix for the various possible transitions. This will not be a refutation of the theory, for one might always chalk it up to the fundamental indeterminacy of the underlying dynamics.

As an example, consider again a theory similar to Bohr's 1913 model that allows no more than one quantum number. As we touched on in §3.3.1.3, $3p$ can transition to $1s$ whereas the other $n = 3$ states cannot (Nasser 2012, 5; see Jitrik and Bunge 2004 for detailed transition amplitudes). So there is failure of tracking. Now suppose an experimenter does not have the tools to prepare the states separately according to azimuthal quantum number (which was probably true at the time). As a result, it would simply appear to such an experimenter that $n = 3$ states *sometimes* transition to $n = 1$, though they often do not. This would make the theory "more indeterministic than it needs to be", but that would not amount to disconfirmation unless a more deterministic set of state assignments can be demonstrated. To be clear, the real theory proposed by Bohr in 1913 did not prohibit one from introducing other quantum numbers to account for the finer distinctions (as Sommerfeld immediately set out to do). But my cases are easier to illustrate with a fictional version of Bohr's theory that allows for no quantum numbers besides n.

In fact, even if a deterministic alternative were found, it would not incline the scientist to abandon the current theory *unless the more fine-grained states could be separately prepared.* Indeed, this is precisely where things stand with modern quantum mechanics. Imagine for a moment that the world is as the de Broglie-Bohm theory claims. That would mean that what standard quantum mechanics calls a single Hilbert state is in fact an equivalence class of a large number of distinct states, corresponding to the same wavefunction but different initial positions. From this point of view, the indeterminacy of quantum mechanics is because it

lumps together several states that *do not co-transition*. What is a matter of fundamental indeterminacy for one theory is thus due to inadequate coarse-graining in the other. However, since one cannot in fact prepare the individual Bohmian states separately, the more deterministic state assignments are of no use. Quantum mechanics might be forever suspect due to its indeterminism, but it will not be *disconfirmed* unless the states that fail to co-transition (if there are such) can be prepared separately. I suggest that this is the real reason why the de Broglie-Bohm theory is unpopular among physicists, and why it is claimed, somewhat confusingly, that it is unverifiable. From a descriptive-ontological point of view, Bohm's theory is as verifiable as its rivals, but from a prescriptive-dynamical view, its state assignment rules are experimentally useless.[13]

To sum up, tracking is sufficient and "almost necessary" for predictive success. In those cases in which tracking is not strictly necessary (i.e. when individual states within a given class cannot be prepared separately and there are no in-class transitions), failure of tracking always at least arouses a suspicion in the scientist that perhaps a more fine-grained set of state assignment rules is called for.[14]

3.3.2.3 Tracking and referential success

I made the case above that tracking cannot be reduced to referential success in the case of Bohr's theory. I will flesh this out in more general terms in this section.

At first glance, one might be tempted to say that tracking is a form of referential success. After all, does tracking not imply that the higher-level states refer to those lower-level states that transition indistinguishably? In other words, are the theoretical terms not picking out certain natural groupings of the underlying states that evolve similarly? The answer is no. There is nothing *natural* or *inherent* about the groupings, because the equivalence classes do not share any property that could be described purely at the level of truth and without

[13]The above arguments apply to cases where tracking fails due to failure to co-transition. Similar considerations would apply to cases in which tracking fails due to difference of magnitude and amplitude.

[14]Note: In the foregoing, I always assumed the mapping from true states to theoretical states to be many-to-one. This is because as long as the mapping is one-to-one or one-to-many (but never many-to-one), and as long as transition rules are read off empirical data as I claim they must be, the theory will be retained. I will leave it to the reader to convince themselves of this. (Hint: for the case of one-to-many mapping, think of gauge freedom.)

reference to the higher-level theory. To see this, first note that successful state assignments are not unique. Consider a deterministic system with four true states $S_1, ..., S_4$ and let the transition rules be $S_1 \rightarrow S_3$ and $S_2 \rightarrow S_4$, and suppose the characteristic magnitudes of the two transitions are roughly equal. Now consider the following three ways of mapping ("lumping") the underlying states into equivalence classes of theoretical states:

1) Assign $\{S_1, S_2\} \leftrightarrow S_T$, $\{S_3, S_4\} \leftrightarrow S'_T$
2) Assign $\{S_1, S_2\} \leftrightarrow S_T$, $\{S_3\} \leftrightarrow S'_T$, $\{S_4\} \leftrightarrow S''_T$
3) Assign $\{S_1\} \leftrightarrow S_T$, $\{S_2\} \leftrightarrow S'_T$, $\{S_3\} \leftrightarrow S''_T$, $\{S_4\} \leftrightarrow S'''_T$

It is easy to verify that the coarse-grainings in (1) and (3) result in a set of state assignments that track (co-transition), whereas the ones in (2) fail to track (co-transition). Now compare the assignment $\{S_1, S_2\} \leftrightarrow S_T$ in (1) to the same assignment in (2). Both assignments lump the same underlying states into the same equivalence class, yet one is successful and the other is not. Clearly, the success of (1) is not due to the fact that S_T "picks out a real property" of the system, or else the same assignment would not fail in (2). The proponent of referential success might try to attribute this to the fact that (2) fails to lump S_3 and S_4 together. Perhaps by failing to put the latter two states in the same equivalence class, (2) fails to represent an important property that they share and thus fails to predict. In other words, perhaps (2) is being too fine-grained for its level. But this cannot be right, because the same fine-grained assignment works perfectly well in (3). In short, the success and failure of state assignments cannot be decided in isolation. It is *the entire set of state assignments* that either tracks or fails to track *as a whole*.

Thus, "the common attribute", if there is any, that is supposedly shared by all the states in the same class cannot be cashed out in terms of any properties that reside entirely at the level of true states. For this would require describing an inherent, theory-independent property that S_1 and S_2 have in common, but no such property is forthcoming: without reference to the theoretical equivalence classes, there is no natural description of S_1 and S_2 that unites them together. In short, predictive success is not a matter of each higher-level state picking out an inherent property of some lower-level states, but a matter of how the entire set of coarse-grained state assignments *as a whole* cooperates with the dynamics.

Similarly, "transitioning indistinguishably" cannot be cashed out without reference to

our resolution limits. For instance, the Schrödinger theory only tracks the true states if one ignores hyper-fine structure, which is due to quantum field theory corrections and cannot be accounted for in Schrödinger's theory. Classical mechanics only tracks natural systems as long as we do not (or do not care to) have the resolution to distinguish the myriad underlying states that it lumps together as the same state. Nature does not recognize these lumps; they only make sense for a being that has a limited resolution limit.

What about structural representation? Can we not say that Bohr's theory represents part of the *structure* of the underlying dynamics? This might seem promising at first, because one of the many definitions of "structure" is *relational properties*, and it might seem that while tracking is not about individual theoretical states representing specific entities or properties, it does have to do with the relations among the theoretical states "mirroring" the relations among the underlying states. My response to this line of reasoning is to say, first of all, that I would not be surprised if out of the plethora of ways "structure" has been cashed out, some of them would fit the bill here. Once a concept is so flexible as "structure" is in philosophy of science, its successful application is hardly impressive. That being said, I am not sure that even the "relational properties" view of structure would apply here, because the only "relations" at play here are the transition rules. That is, one must define a relation that obtains between two states when one transitions to the other. But what about all the other relational properties that one could define among the underlying states? Take for instance the most obvious one: the relative differences in their characteristic magnitudes (e.g. difference of binding energy between various atomic states). As we have seen, Bohr's theory does not represent the "structure" of these relations, unless one abstracts a coarse-grained structure from it, which can only be defined by taking the transitions into account. Now, perhaps the proponent of referential success is willing to narrow "structure" down to something like "transitional structure", and perhaps one could then say that Bohr's theory represents this structure in some way, but at that point the structuralist would simply be saying precisely what I have been saying, except in a convoluted and confusing language. The language better suited for this context is that of dynamical systems, which has been my instrument throughout.

As we have seen, tracking is irreducible to referential success because it is the result of a

confluence of four factors: the way the underlying states are mapped to theoretical states, the underlying dynamics, our experimental resolution limits, and the (un)availability of various preparation procedures. "Referential success", if at all relevant, at best captures something about the first component, namely the mapping between true and theoretical states, but it leaves out all the other crucial factors.

The unknown admixture of the above factors means that tracking is a *referentially opaque* relationship, because the mapping cannot be "inverted" to find the true states unless the other factors can be eliminated from the mix. And this brings me to the last but not least difference between tracking and referential success.

3.3.2.4 Tracking and referential transparency

I have used the phrase "referential transparency" on several occasions. It is time to clarify the concept. One way to formulate the idea of referential transparency is as follows: certain components of successful theories give us direct knowledge of the ontology of the world. In other words, if a theory is related to reality in a referentially transparent manner, then one can "read off" the ontology of the world as soon as one knows which component of the theory to believe. To put it in yet another form, referential transparency means that identifying the essential content requires *demarcation*, but never *reformulation* or *reconstruction*: the privileged components are simply "sitting there" somewhere in the theory, ready-made for us, and all we need to do is carefully draw a line around them to separate the wheat from the chaff.[15] Formally speaking, if one thinks of theory-world relations as mappings that take us from elements of reality to theoretical terms (such as the map that takes us from Schrödinger states to Bohr states), referential transparency means that this mapping is a "partially invertible" one.

Referential transparency also implies that identifying the essential content allows one to predict (parts of) the ontology of *future theories*.[16] This is because referential transparency

[15]Note that referential transparency does *not* mean separating the wheat from the chaff would be easy.

[16]Peters (2012, 139) and Leconte (2017, 3263) both mention this as a problem for current accounts of essential content, for it would imply that we could predict the future of our current best theories, which is presumably absurd. I am sympathetic to this objection, but my criticisms below will be based on other considerations.

implies that one must expect the essential content to survive revolutionary theory change more-or-less intact. For if a theoretical constituent directly corresponds to an element of reality, no alteration in its content is to be expected when it reappears in future theories: it will simply be "carried over" to the superseding theory. Thus, adherence to the idea of referential transparency sends one searching specifically for theoretical constituents that have been *retained*. Conversely, the assumption that essential content is always retained in future theories more-or-less intact implies that those posits transparently refer to some element of reality, for otherwise they would be expected to be substantially reformulated in future theories. Therefore, referential transparency and retention mutually imply each other.

> *Referential transparency:* The ontology of the world (and that of future theories) can be directly read off the essential content, which must simply be demarcated (but not reformulated) and which will be retained in future theories more-or-less intact.

The long-standing notions of "representation", and "reference" are the canonical exemplifications of referentially transparent theory-world relations.[17]

Bohr's theory is doubly referentially opaque. First of all, as said above, its entities, properties, and structures are not reflected in reality as we know it. There are no point particles resembling tiny grains of dust with mass and charge, and there are no entities that orbit the nucleus in elliptical paths, etc. But the second – and in my opinion more interesting – sense in which Bohr's theory is referentially opaque is that *even its states* do not transparently represent or refer to the true states of the system, despite the fact that we have identified the state assignments as the essential content. Put differently, although Bohr's state assignments as a whole "latch on to" certain aspects of unobservable reality, the mere fact that an individual Bohr state has been correctly assigned to a system tells us very little about the actual state that the system is in at the time of assignment.

More generally, from the mere fact that a theory successfully tracks a system one can infer very little about the true entities, true properties, true structures, or even true *states* of the system. That is, as pointed out by Hertz, nothing except that they transition in such a way that is compatible with whatever unknown mapping our theory is imposing

[17] And insofar as one accepts Laudan's (1981, 33) argument – and it is typically accepted by both sides – that the central terms of a truthlike theory must refer, *truthlikeness* also implies referential transparency.

on them. Further, the fact that a theory tracks a system does not imply that anything from the theory will be recognizably *retained* in subsequent theories. Of course tracking *does* mean that there is a robust, systematic relationship between the unobservables of the theory and those of reality, and this systematic relationship explains the empirical success of the theory. Moreover, this explanation goes beyond antirealist explanations, such as van Fraassen (1980, 40) and Wray's (2007; 2010) evolutionary explanation, Laudan's (1984) methodological explanation, Fine (1986) and Lyons's (2002, 78; 2003) "as if" explanation, and Stanford's (2000) predictive similarity. The former two types of explanation assume no relationship between theory and reality at all, while the latter two assume a relationship between theory and world, but a rather superficial one that is not systematic at the level of unobservables.

Therefore, the prescriptive-dynamical view can be considered a form of selective realism, but a rather novel one. It is *selective* in that it urges us to ignore the ontology (entities, properties, and structures) and only take the state assignments seriously as the essential content of the theory. And it is *realist* in the sense that it supports the idea of a systematic relationship between theoretical state assignments and the true dynamics of the system, and claims that the theory's predictive success can be explained non-miraculously through this systematic connection to reality. However, the prescriptive-dynamical view is *not* a realist position insofar as "realism" is taken to involve the claim that the systematic theory-world relation must be *referentially transparent*, i.e. that the essential content must be *approximately true*, or that they *refer* to or *represent* objects in reality. Therefore, the prescriptive-dynamical view provides a middle ground between the polar positions of realism and antirealism. As such, whether the prescriptive-dynamical view is "realist" or "antirealist" might be to some extent a matter of terminology. In any case, the exact relationship between my view and the realism / antirealism divide is the topic of Chapter 4. For this chapter, I will leave it at that.

3.4 Assessing the Existing Accounts of Essential Content

In this section, I will firstly provide a general taxonomy and list of accounts that currently exist, and secondly highlight the extent to which the (implicit) assumption of referential transparency has affected the cogency and coherence of these accounts. Recall that referential transparency implies and is implied by the assumption that the essential content must be *retained* in future theories more-or-less intact. I will argue that most existing accounts are unsatisfactory because of their insistence on retention.

But in fact my objection applies to *both proponents and critics of selective confirmation*. I shall argue that the critics also share the (implicit) assumption of referential transparency, and that this assumption is the reason why these critics miss the mark about selective confirmation in general. To be clear, I find many of their critiques effective against the extant accounts of selectivism, but I would argue that many of the same criticisms are ineffective against a version of selectivism that relaxes the assumption of referential transparency, such as the one I have offered.

Note: In the interest of brevity, I will not provide an exposition of the existing accounts in this section. But since I am making substantial claims about what each account presupposes, I must provide arguments and/or textual evidence. I have done this in Appendix C.

3.4.1 The presumption of referential transparency among the proponents of selective confirmation

3.4.1.1 Synchronic indispensability accounts

The first and most natural way one might choose to tackle the issue of selective confirmation is through the idea of *indispensability* or *ineliminability*. The essential or working posits are those constituents that cannot be eliminated without damaging the predictive inferences of the theory. In a way, that is the very definition of essential content. However, this definition can also be made into a *criterion* for choosing essential content, if one assumes that dispensable posits can be directly *detected* as such by examining the flow of concrete predictive derivations. The idea is that one can start eliminating various components of a

given derivation until one reaches a bare-bone argument from which no further component can be eliminated without undermining the conclusion. Whatever is left in this bare-bone argument is then taken as the essential / working part of the theory.

The idea has been attributed, among others, to the original founders of selectivism, namely Kitcher (1993, 143-149) and Psillos (1999, 110). The view attributed to Kitcher entails *retention of entities and/or properties*, while that of Psillos presupposes *retention of hypotheses*. Vickers (2013) sketches an algorithm for finding the essential content that also assumes *retention of hypotheses*. And finally, Saatsi (2005) and Egg (2016) have offered synchronic indispensability accounts which require *retention of properties*. (See Appendix C for details.)

Specific considerations may be and have been raised against each of the above accounts (see e.g. Lyons 2006; Stanford 2006; 2009; Chang 2003; Peters 2012; 2014). But I would like to focus on a *general* problem that I believe affects all of the views above. This problem stems from their common assumption of *retention*, which as said above is equivalent to referential transparency. Synchronic indispensability accounts claim that there is a hidden gem in every theory that will be preserved through all manner of theory change. This may be an entity, a property, or a hypothesis, which will in the future find itself within a radically different ontological environment but nevertheless manage to survive. Theory change involves ontological implants, as it were. For instance, the assumption of an elastic solid being the medium of light did not survive theory change, but certain properties of ether did, for instance that it produces transverse vibrations.

What is rarely disucssed is whether and how such ontological implants are possible. If a property is to be transferred to the ontological framework of a successor theory, the host theory's nexus of inter-related entities and properties must be "receptive" to the implant. For instance, many properties of the electron, such as charge, have survived many episodes of violent theory change since Thomson's experiments despite the fact that our understanding of the entity as a whole has radically changed (cf. Norton and Bain 2004). This is because the new ontologies to which properties such as charge were transferred contained entities that were capable of carrying charge.

The problem is that this is not always or even often the case. For any given entity or

property, it is often the case that if one waits long enough, along comes a successor theory that is simply ontologically too different to receive that particular entity or property without jeopardizing its internal coherence. For instance, the concept of *force*, taken seriously, cannot exist in general relativity without making nonsense of the latter theory. By "taken seriously" I mean force in its full glory as it appears in classical physics: an entity of its own that emanates from an object and acts on another to deflect its otherwise inertial path. No such thing exists in general relativity, and more importantly, no such thing *can* exist in this theory. For according to general relativity, all objects at all times follow their geodesics. Nothing ever *acts* on an object and nothing ever "deflects" the object from its geodesic path. Further, the gravitational effects that we recognize as such, for example the earth going around the sun, are *not* the result of anything emanating from the sun, but rather due to the local spacetime curvature that the earth feels. To be sure, this curvature is influenced by the presence of the sun, but there is no vectorial quantity whose point of impact is on earth with its direction pointing towards the sun's center of mass and its magnitude proportional to earth's acceleration. Nor is there a gravitational potential, as the very concept of potential energy is often ill-defined in general relativity.

It seems to me that situations akin to gravitational force in GR are more common in the history of science that those like charge in quantum mechanics, and they become more likely as more revoultions take place in the theoretical lineage. This should not strike one as surprising given that theories are often substantially *reformulated* and *reinterpreted* in light of subsequent developments. What made an old theory successful often has to be cashed out in terminology radically different from the language used to formulate it originally. What is surprising is that despite these truisms about theory change, philosophers have universally insisted on the rigid concept of retention.

Indeed, I believe the fate of ether theories presents one such case of substantial reformulation. As said above, selective realists have repeatedly claimed that the entity called ether did not survive but some of its properties did. But since a property needs an entity as a vessel to carry it, the property in question must now be carried by an entity other than ether in the new theory. Now, this new entity must be very different from ether: that is presumably why we are inclined to say that ether itself was not preserved. But the more dif-

ferent this new entity is, the less reason we have to hope that this new entity can accept the properties of ether without destablizing the new ontological framework. Take the property of vibration, for instance. If we take the meaning of "vibration" seriously and in the way it was intended in ether theory, then there cannot be vibrations in the modern theory, as there is nothing to vibrate. The undulations in the electromagnetic field values in different spatiotemporal locations are *not* vibrations. The very concept of vibration implies material bodies that are rapidly and repeatedly moving to and fro around an equilibrium position. *No such thing is happening according to modern electromagnetism.* Sure, E and B field values vascilate in space in such a manner that if one fixed the frame of reference (since E and B trade magnitudes upon change of frame) and went on to interpret each maximum as a material body being the fartherst from its mechanical equilibrium and each minimum as the body returning to its equilibrium, then one could present the vascilations as "vibrations". But apart from the fact that the fixing of reference frame would be totally arbitrary and unphysical, the peaks of E and B fields are *not* bodies stretched farthest from the equilibrium for there is no "restorative force" inclining them to return, and when the field values become zero, that is *not* a case of the body returning to its resting place for the value zero is not special for E/B fields whereas the equilibrium point is special for a harmonic oscillator. Moreover, vibration implies that *the same body* is going back and forth, but when the E/B field acquires a non-zero value again, having momentarily vanished at some point along the wave front, there is no sense in which it is "the same" field acquiring non-zero value again.

Stanford (2006, 171 ff.) has famously objected to Kitcher and Psillos's treatment of the ether case in part by arguing that such ontological selectivism would have been *unintelligible* to contemporaneous scientists. He produces quotes from Maxwell, for example, to show that he did in fact consider vibrations without a medium and judged it to be nonsensical. Stanford brings this up as a way of showing that such butchering of ontology does not make sense *except in hindsight*. But as I see it, hindsight is irrelevant to this problem: such ontological butchering does not make sense, period. It did not make sense during Maxwell's time, and it does not make sense now. That is an analytic fact, and has nothing to do with hindsight or historical context. There can be no vibrations without something that vibrates. End of story. Thus, the problem with these accounts is not that they are retrospective, but that

they are retentionist.

To be clear, I am not denying that there is some kind of abstract similarity between vibrations of an infinite system of oscillating molecules in an elastic solid and the undulations of continuous E and B fields[18], but my point is that if one is being metaphysically honest, ether's property of vibration, as such, has not been preserved; it has rather morphed into something that it was not. In a way, there is often a confession to the same effect in the strange language that realists use to talk about retention. Having claimed that entities, properties, structures, or hypotheses of past theories have "survived" theory change, the realist immediately goes on to say that successor theories "converge" or "degenerate" to older theories as "limiting cases" (see, e.g. Peters 2012, 45 ff.; Votsis 2011b, 4). But converging to something in the limit is not the same as containing or retaining it. It is true that general relativity goes over to Newtonian gravity if we assume slow motion of bodies as well as a static and nearly flat metric, but that is a far cry from saying that gravitational forces are *retained* in general relativity, which as I argued above, they certainly are not. If one insisted on the language of retention in such cases, retention would be far too easy to come by to have any significance (see Lyons 2016, 103 for similar complaints about the idea of similarity "in the limit").

In summary, it is often impossible to rip away a given posit from its home ontology and force it to migrate into a new web of concepts in which it has no place. Usually, if the displaced posit is to maintain its essence, the new framework must be awfully similar to the old one, rendering the case of theory *change* less interesting. If the new framework is not similar to the old one, on the other hand, then the old posit will most likely render it incoherent. Cordero has expressed this elegantly: "Scientific theories are tight constructs, and breaking them into parts is generally not doable... ." (2011a, 24) I conclude that synchronic accounts of essential content that rely on a notion of retention (of entities, properties, or hypotheses) are unlikely to succeed.

[18] A structuralist would call this a "structural similiarity", but I believe there is a more meaningful way to describe this similarity, namely in terms of state assignments tracking each other. See below.

3.4.1.2 Diachronic indispensability accounts

The indispensability accounts we reviewed above are *synchronic* in the sense that they claim to identify the essential content of a theory *while the theory is still in place*. More recently, many *diachronic* accounts of indispensability have been proposed. The idea behind these accounts is that one cannot know the essential content of a theory until it has been thoroughly dissected, modified, and/or replaced in light of future investigation. Unlike Psillos (1999, 112), who claims that contemporaneous scientists are often reliable judges of indispensability, the proponents of diachronic accounts typically argue that scientists who know what is (in)essential about a theory are those who investigate ways to *modify or replace* the theory.

Note that although diachronic accounts are "retrospective" in the sense that they suggest we find the essential content of a theory by examining subsequent theories, they are not vacuous or question-begging in the way that Stanford worries all indispensability accounts will be (see below). The reason is that diachronic accounts do not ask us to consider what has been retained in *current theories*, but typically in *immediately succeeding theories*. Diachronic accounts are therefore fully testable (assuming they provide unambiguous selection criteria) and hence cannot be vacuous. To test a diachronic account of essential content, simply apply the diachronic recipe to identify the essential content of a theory by examining its immediate successors, and then see if the same content has been preserved in our current best theories. If it has, then the diachronic account gets a boost; if it has not, the account gets refuted.

Harker (2010; 2013) proposes a diachronic account that adheres to *retention of entities and hypotheses*. The same goes for Cordero (2011a; 2011b). Peters (2012) has suggested a way of identifying essential content that may also be categorized as a diachronic account, and it presupposes *retention of entities and/or structures*.[19] (See Appendix C for details.)

As before, I am interested in a very general problem that all of the above-mentioned accounts have: namely their insistence on "retention" and "preservation". In the previous section, I argued that synchronic accounts must play fast and loose with the latter concepts,

[19]Other diachronic accounts include Badino's (2016) criterion of long-term entrenchment in "reliable symbolic practices", Onishi's (2017) reliance on scientists' long-term consensus, and Sakellariou's (2011) account of theory maturation.

primarily because they overlook the fact that a theory must often be substantially *reformulated* and *reinterpreted* in light of future theory before its essential content can be identified. Diachronic accounts are more promising in this regard, since they claim that one must wait for future theories to extract the essential content of a successful theory. This seems to suggest awareness of the need for reformulation. Unfortunately, such awareness has not been expressed so far.

If the theory need not be couched in radically new language in order to identify its essential content, then why is it that, in order to find it, we must wait for a tortuous process of distillation that according to the diachronic selectivist might take centuries? Why can we not simply look at the old theory while it is still practiced and find the essential content? Stanford (2009, 385) challenges the selectivist to explain why contemporaneous scientists ever believed anything more than the working posits unless, as he argues, it is impossible to discover such posits except in retrospect. Cordero (2011a; 2011b) provides a possible answer: because there is often too much metaphysical prejudice among contemporaneous scientists. But Cordero's answer is not satisfactory. It is not as though the successor theory's only role is to disabuse us of prejudice, and as soon as that is done, suddenly our eyes are opened to the obvious working posits in front of us that we had so stubbornly refused to see. Rather, the successor theory allows us to see the working posits because, along with disabusing us of prejudice, it also reformulates the old theory in radically new language. Thus I believe that there is an answer better than either Stanford's or Cordero's to the question raised by the former: contemporaneous scientists often fail to see the essential content for what it is *because the theory often needs to be substantially reformulated* before this can be done. I therefore find Stranford's objection forceful against the diachronic retentionist accounts such as above, but not against selectivism in general.

3.4.1.3 Structural realism

Structural realism is also considered a form of selective realism, as it urges us to ignore the "content" of a successful theory and focus on its "structure" instead. This view was brought to the contemporary scene by Worrall (1989), in a seminal paper in which he argued that

structural realism would address the forces of both No Miracles and Pessimistic Induction. Worrall's idea struck the fancy of many philosophers of science, who have since advocated various versions of it (see Ladyman and Ross 2007; French and Ladyman 2003; 2011; French 2006; Votsis 2011a).

Our concern is specifically with structuralist views regarding selective confirmation. These include Worrall (1989; 1994), Ladyman (2011), and Votsis (2007; 2011a; 2011b). All three selectivists presuppose *retention of structure*. (See Appendix C for details.)

Apart from specific objections to specific versions of structural realism (e.g. the Newman objection against Ramsification), there are two major objections against structural realism in general: i) that the notion of "structure" is too vague to be contentful, and ii) that structural realists have not provided a reason to think that structures are success-generators.[20] I suggest that both of these problems can be traced back to the assumption of retention / referential transparency.

The vagueness objection is by now fairly common in the literature. Harker (2010, 198), for instance, complains that the definition of structure is "disconcertingly vague"; Lyons (2016, 96) argues that structuralist responses to counterexamples have increasingly weakened the notion of "structure" so much as to make it "vacuous". See also Vickers (2016, 7), Saatsi (2017, 3238-3239), and van Fraassen (2006, 290) for similar considerations. The situation is so bad that Votsis (2011b, fn. 2), a robust advocate of structural realism himself, admits that at a workshop featuring most structural realists, the participants could not find one thing that their various notions of structure had in common. This is not at all surprising, given that upon a quick survey of the literature, one finds "structure" variously identified with: the mathematical equations themselves (Worrall 1989, 158), Ramsey sentences (Worrall 2007), set-theoretic / model-theoretic objects (French 2011; French and Ladyman 2011), category-theoretic objects (Landry 1999; Bain 2013), frame-theoretic objects (Votsis and Schurz 2012), group-theoretic symmetries (French 1999), whatever is

[20]In Ch. 2, I also objected to structuralism on the ground that theories do not have fixed structures due to their open-ended nature. It has been shown for instance that the structure of the phase space depends in part on the specific Hamiltonian being used, and according to my view Hamiltonians encode transition rules, which are local and empirical (LEMPs) and hence open-ended. Cf. Lyons 2016, 101 for a similar objection in the context of general relativity. Note that Lyons seems to view this open-endedness as a peculiar feature of general relativity, while I have argued it is quite generic of physical theories.

invariant under coordinate transformations (van Fraassen 2006, 291; North 2009), second- and higher-order properties (Maxwell 1970, 18), relations[21] among objects (Ladyman 2007; French and Krause 2006, 172), relations among phenomena (Ladyman 2011, 98-99), modal relations (Ladyman 1998; 2004; French and Ladyman 2003, 46), and a combination of state space, dynamics, and symmetries (Reutsche 2002, 200; Bain [manuscript], 24). What one never finds is whether and how any of these definitions are equivalent or even compatible. In his attempt to clarify the notion of structure, for instance, Votsis (2007, e.g. 58-59, 62-63) ironically moves swiftly between many of the above definitions, often within the same paragraph, without commenting on their possible equivalence. As Lyons (2016, 96) has put it: "there are more variants of structuralism than there are structuralists"!

The vagueness problem is rooted in the structural realist's insistence on retention. Since the latter commits one to finding something that can readily be recognized in both theories as *the same structure*, and since as I argued above such things are very difficult to come by, the structuralist's strategy is often to look at the two theories' central equations and find *something* that looks similar between the two, and claim that that something is the "structure" of the theory. For instance, consider two supposedly lucid examples Ladyman and Ross (2007, 94-95) introduce in two consecutive paragraphs as a way to ease the reader into the idea of structural realism: the first is a comparison between Galileo and Lorentz transformations[22], and the second between Newton's $F = ma$ and a similar-looking formula that can be produced from the principles of quantum mechanics using Ehrenfest's theorem. It is not clear under what definition of structure these examples are both cases of structural similarity. After all, the two pairs of equations are of radically different nature.

For one thing, one pair contains equations governing coordinate transformations which say nothing about dynamics, while the other contains dynamical differential equations pertaining to how objects move. The only thing the two pairs have in common is that both are mathematical equations that look similar. This would perhaps be acceptable for those who define structure as bare mathematical equations (never mind that those people are more-or-less extinct these days, see Votsis 2007), but Ladyman and Ross are certainly not among

[21] Even the concept of "relations" is murky in this literature, as it is sometimes meant to exclude monadic properties and others to include them. Cf. Votsis 2007, fn. 6

[22] Votsis 2011b, 6 also uses this example in service of structuralism.

those people.

Secondly, the manner in which the two equations are similar is quite different. In the case of Galileo vs. Lorentz transformations, the similarity consists in the fact that the two sets of equations look the same except for a multiplicative factor of $\gamma = \frac{1}{\sqrt{1-\frac{v^2}{c^2}}}$. In the case of classical vs. quantum mechanics, however, the similarity is that one can convert one equation into the other by substituting symbols representing the expectation value of the position operator with symbols representing classical position. And as we saw in §3.2.1.2, Vickers finds structural similarity between two equations that differ by an *additive* term that becomes small in the limit. One gets the impression that all there is to the structuralist's claim is an excited exclamation: "Look at them! See how similar they look!" As Stanford (2009, 387) puts it in another context, the structuralist seems to be happy as long as "something [...] somehow [...] somewhere" is preserved. This is of course to be expected if one's ultimate goal is to find *something*, anything, that looks similar in the two theories, rather than fixing beforehand what one is looking for and then test to see if they are indeed similar.

The brings me to the second problem with structural realism, namely that it does not explain how structures bring about predictive success. Following Psillos's (1999, 146-161) objections to structural realism, Harker (2010, 197) remarks: "A further argument is needed to the effect that, if a theory correctly describes structural aspects of the world, then it can be successful... ." After all, the aim of the selectivist is to identify those posits that are *responsible* for success, not simply to find something or another that has been retained (Harker 2013, 84-85). Pashby (2012, 470) points out that it is not enough to point to structural continuities through theory change, because there are often structural *discontinuities* as well. One has no reason to point to the continuities as the source of success unless the discontinuities have also been examined and shown to be irrelevant. Saatsi (2009, 333; 2017, 3238-3239) makes similar remarks.

The two examples provided by Ladyman and Ross signal the root of this problem as well: as said above, one pair of equations designates coordinate transformations while the other provides dynamical laws and, moreover, in the former case the similarity is due to difference by a multiplicative factor, while in the latter it is due to the existence of a substitution procedure that turns one into the other. One needs to hear a story about how each of these

equations brings about success and why what they have "in common" is indeed the part that generates success. On the contrary, the alarming truth is that in each case study, structural realists all agree that the two theories have a structure in common, but they cannot agree on what the two theories have in common (cf. Votsis 2011b, 3). Once again, this seems to be due to the fact that the structuralist insists on finding something that has been *retained*, not something that is resopnsible for success. As I argued above, essential content is typically not retained. Nevertheless it is not surprising to find several strings of mathematical symbols in some formulations of two empirically overlapping theories that look kind of similar to / continuous with each other (cf. Pashby 2012, 465; Vickers 2016, 7). These two facts lead to the structuralist identifying a different aspect as the "structure" every time. I therefore agree with Lyons's diagnosis: "the structuralist's need for the retention of structure compels the structuralist toward increasingly vacuous conceptions of 'structure'." (Lyons 2016, 96)

As a concluding remark, I would venture to guess that in those cases where the structuralist has successfully identified the essential content, the "structural similarity" is better described as a compatibility of state assignment rules. Indeed, Wallace (2012) has shown this in the case of Ladyman and Ross's example of classical vs. quantum mechanics. The situation with Fresnel and Maxwell is also a case in point. It is true that the success-generators of Fresnel's theory were his differential equations which carried over to Maxwell's theory. But since differential equations are input-output machines that take initial states to final ones (see §3.3.2), it seems much more helpful to say that Fresnel's theory was successful because its state assignments, as captured in the differential equations, track those of the modern theory. Similarly, I would argue that the state assignments of phlogiston and caloric theories track certain transitions seen in processes of oxydation / reduction and heat transfer, respectively. But a satisfactory treatment of the cases of caloric and phlogiston in the prescriptive-dynamical view is beyond this chapter.

3.4.2 The presumption of referential transparency among the critics of selective confirmation

Critics of selective realism are just as guilty of presupposing retention / referential transparency as its proponents. In this section, I will argue that this limits the scope of validity of such criticisms.

Lyons (2002; 2006; 2009; 2016) marshals many interesting and potentially devastating counter-examples to selectivism, but all of his counter-examples assume that selectivism presupposes *retention of hypotheses*. Stanford (2006; 2009), one of the fiercest critics of selectivism, similarly assumes the centrality of *retention of entities* in selectivism. (See Appendix C for details.)

I suggest that Lyons and Stanford's criticisms are valid *but only against versions of selectivism that adhere to referential transparency*. Consider Stanford first. He objects that selective realist views are bound to be retrospective (and thus uninteresting and/or totally uninformative), as they require moves that would have been as much as unintelligible to contemporaneous scientists. But the fact that identifying the essential content prospectively is often difficult and the fact that the resulting selective picture is often incoherent / unintelligible are both rooted in the referential opacity of physical theories. Take the issue of retrospectiveness. Because of referential opacity, the theory must be substantially reformulated before its essential content becomes apparent, and it may be hard to find a footing for this reformulation without help from successor theories. As we saw above, this does not necessarily mean drawing on *current* theories, as Stanford claims. Retrospective does not automatically mean question-begging or vacuous.[23] Therefore, taken as a criticism of all retrospective criteria (taking "retrospective" to include diachronic accounts), Stanford's objection is a straw man. But there is a valid point hidden in Stanford's apparent overstatement: the essential content is not just some subset of the theory's posits as formulated; it is often buried deep within the theory. This is the root of Stanford's retrospectiveness objection. Similarly in the case of the uninformativeness objection, which is rooted in Stanford's claim about the unintelligibility of the proposals. In §3.4.1.1 above, I argued extensively that

[23]This point has been made in the literature: cf. Saatsi 2009, 361-362; Saatsi 2017; Sakellariou 2011, 121; Votsis 2011a, 1230-1231; Peters 2012, 135 ff.; Onishi 2017, 8.

this unintelligibility is due to insistence on retention, which results in the "butchering" of the theory's ontology. This is because of referential opacity: the fact that the theory needs substantial reformulation to reveal its essential content means that the entities, properties, etc. featured in the original formulation need not have any straightforward connection to elements of reality. This means that one cannot hope to find the essential content by simply isolating one component out of a theory and claiming that it represents something real that will survive in future ontologies. Insisting on doing so often gives nonsensical results, not just to contemporaneous scientists, but to anyone who adheres to the law of non-contradiction. This is the root of Stanford's uninformativeness objection.

Next consider Lyons. Lyons argues that it is impossible to find anything truthlike among the premises that Kepler used to derive his area law, or those used by Laplace and Haüy to derive the fixity of the rate of expansion of dilute gases. Nor does it seem that any of the weaker implications of these premises can be claimed to be even approximately true. Now, I do not contest this claim, but it seems to me that this is devastating to the selectivist *only if one insists on examining each premise individually.* That is to say, if the procedure is to choose one premise at a time and inquire about its essential content and then move on to the next premise, one might well come back empty-handed. But this procedure is only appropriate for someone who subscribes to referential transparency. On the contrary, as I argued above, the theory might need substantial reformulation before its essence can be brought out, and *this reformulation need not contain any of the premises in their original form.* As a familiar example, consider how the reformulation of Lagrange's variational theory as a dynamical system does not contain the principle of least action. Lyons's cases are only knock-down arguments if it has been shown that *there is no reformulation of the theory* that contains respectable essential content. As soon as one realizes this, Lyons's objection becomes much less conclusive. I do not pretend that this solves Lyons's cases, but my point is that Lyons's arguments have a loophole that goes back to the (implicit) assumption of retention / referential transparency. Therefore, the possibility of reformulations that could save referentially opaque accounts of selectivism must be explored before one can pass a verdict about these case studies.[24]

[24]That being said, it is possible that some historical examples may just be rare anomalies that were

As such, Lyons and Stanford's are legitimate objections to every account reviewed above, but they are not legitimate objections to selectivism in general, for it is possible to be a selective realist without endorsing retention / referential transparency. Of course Lyons and Stanford might retort that they are not to blame for this assumption because they are evaluating positions that make that assumption. But in that case they should not present their arguments as if they are general.

3.5 Conclusion

I have argued that selectivism (i.e. the belief that the success of a scientific theory is due to an essential subset of its posits) is motivated independently of the realism debate, both through general arguments and through a case study of Bohr's atomic theory. More specifically, I have argued that the essential content of dynamical theories (such as Bohr's and most any other theory in physics) consists of their state assignment rules. Finally, I have argued that the appropriate form of connection between a theory's state assignments and the true states of the system is captured in the tracking relationship. Unlike other accounts of theory-world relations, tracking is referentially opaque and does not require retention of essential content in future successful theories.

successful due to luck. Consider for instance Lyons's (2006, 551-552) example of the Titius-Bode law, which made novel predictions about the position of Uranus and some other celestial bodies, but which from the contemporary point of view is an arbitrary constraint on initial conditions, which are unconstrained in modern theory. It seems to me that even the antirealist cannot explain the success of this formula beyond luck.

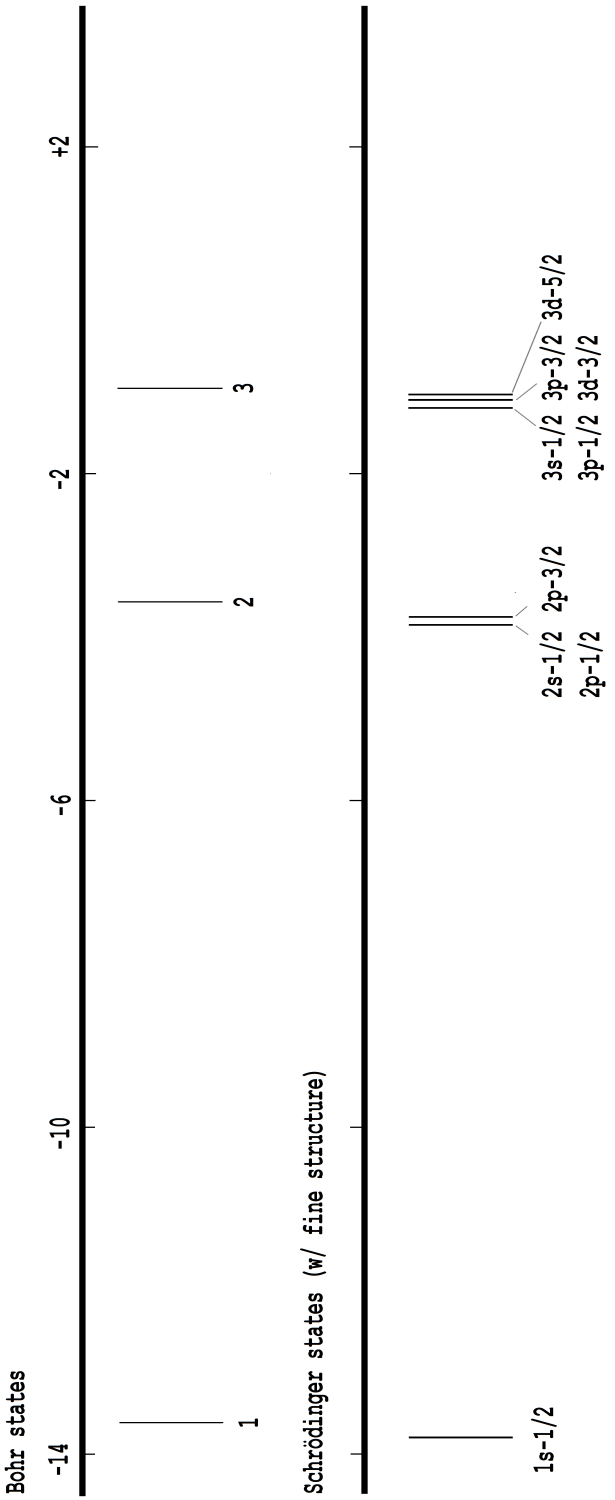

Figure 1: Bohr vs. Schrödinger for hydrogen (fine structure exaggerated)

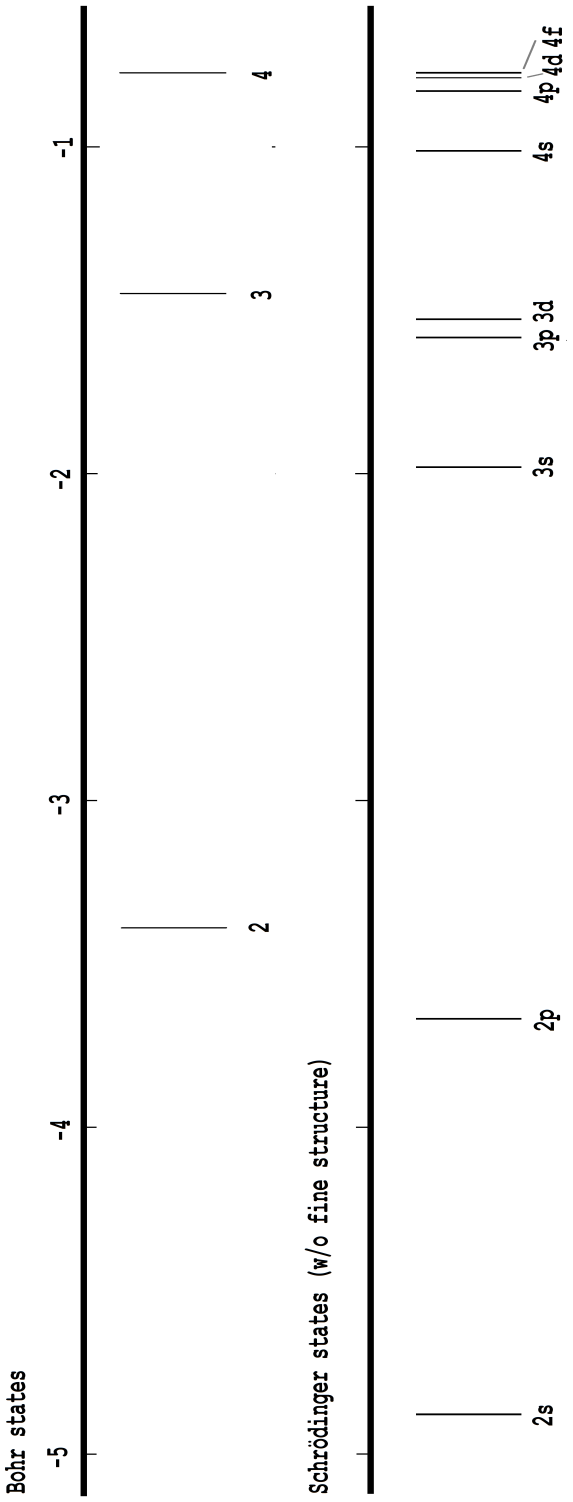

Figure 2: Bohr vs. Schrödinger for orthohelium (fine structure suppressed)

4.0 Chapter 4: Escaping the Fundamental Dichotomy of Scientific Realism

4.1 Introduction

My aim in this chapter is to provide an explanation of the success of physical theories from a dynamical systems standpoint.[1] As such, my focus is on dynamical theories and theories that can be given a dynamical formulation, which includes virtually every predictive theory in physics as well as theories outside of physics which are governed by differential or integral equations. I will not argue for extending my thesis beyond this realm.

My explanation will be both realist and antirealist, or depending on one's taste, neither realist nor antirealist. It is realist in the sense that it explains the success of a dynamical theory through intricate and robust connections between the theory's theoretical terms and unobservable reality, but it is antirealist in the sense that it does not require any sort of referential transparency of theoretical terms (to be clarified below).

In Section 4.2, I will introduce a distinction between robust and unrobust explanations of success and outline their use. In Section 4.3, I will elaborate on the notion of "referential transparency", and go on to highlight a widespread assumption which I call the Fundamental Dichotomy, according to which referentially transparent explanations are the only robust explanations of success. I will provide textual evidence for my claim that the Fundamental Dichotomy has been assumed by realists and antirealists alike and has therefore dominated the realism literature. I shall then proceed to provide criticisms of the Fundamental Dichotomy and the sketch of a third alternative called the "middle path". In light of this, Section 4.4 presents a Modified No Miracles argument that is liberated from the Fundamental Dichotomy and naturally opens up the middle path. Finally, Section 4.5 is devoted to fulfilling the promise of the middle path by demonstrating robust yet referentially opaque relationships between dynamical theories and unobservable reality under the rubric of "tracking".

[1] A dynamical system consists of a set of state assignments and a set of transition rules among those states, provided the transition operator satisfies certain basic conditions. See (Giunti [2006]) for elaboration.

4.2 Robust and Unrobust Explanations of Success

What explains the impressive, unparalleled success of science? One finds two types of explanation for the success of scientific theories in the literature, which I shall call robust and unrobust explanations, respectively. The former group of explanations appeal to particular intrinsic features of the theory's unobservable elements and their relationships to reality in order to explain why the theory works. The latter group of explanations, on the other hand, bracket any deep relationship the theory might have to unobservable reality, and instead appeal to the methodology through which such adequate theories are constructed and selected. Metaphorically speaking, if theories were extremely successful athletes, a robust explanation of their success would appeal to muscle mass, lung capacity, bone structure, and so on, and how these physiological features relate to the objective demands of the task; an unrobust explanation, on the other hand, might suffice it to point out that the athletes have gone through the federation's stringent and competitive program that has selected the most successful ones over time.[2]

Let us define these terms more precisely. The notion of "robustness" is inspired by the invariance-based account of natural laws, which as Woodward ([2018]) has recently characterized,

> links laws of nature to invariance (aka stability, robustness). Laws are generalizations about repeatable relationships that are invariant over variations in initial and other sorts of conditions, at least within an appropriate range of such variations—invariant in the sense that laws will or would continue to hold under such variations. Alternatively, laws are generalizations that exhibit a certain sort of independence from initial conditions [. . .] (Woodward [2018], p. 158)[3]

Woodward draws on some of Wigner's ([1979]) musings on physical theories, according to which the physicist's central strategy is to separate out certain aspects of reality as initial conditions and the rest as invariant regularities. This is in perfect agreement with a dynamical systems framework, for the latter takes the essential content of a theory to be state assignment rules, which in effect tell us what variables to keep track of as the initial and

[2]This example is due to Leplin ([1997], p. 9).

[3]See also (Woodward [2003]; [2013]); (Mitchell [2000]); (Lange [2009]) for other formulations of the invariance-based account of laws.

boundary conditions (aka states), and what variables to record as transitions rules among such states.

Although my concern in this chapter is with explanations of scientific success, not the nature of laws, the notion of robustness is pivotal to my arguments below. For what is impressive—or "miraculous" if you will—about the success of mature scientific theories is not their success in making particular predictions, but rather the robustness of this success. A false theory might get lucky with a series of predictions made under a specific set of conditions, which in itself need not require much of an explanation beyond coincidence or intentional design (example below); but if and when a theory continues to show resilience in making correct predictions across a tremendously wide range of initial, boundary, and external conditions, that would be surprising unless the theory had in some sense "latched on to something". Therefore, I urge, the challenge posed by the No Miracles argument is not to explain success per se, but to explain the robustness of success in the case of mature scientific theories.

Using the notion of robust success, I would like to define a notion of robust explanation of success, which is simply an explanation that actually explains the robustness of the theory's success. To be robust, the explanation must first and foremost be intrinsic, that is it must require mentioning at least some of the unobservable content of the theory as part of the explanation of its success. Secondly, to be a robust explanation, the account must establish a robust relationship between theoretical terms and unobservable reality that holds up under a range of systems and conditions.

> Robust explanation of success: An intrinsic explanation that establishes a robust formulaic relationship between theoretical terms and unobservable reality.

Let me elucidate the above definitions through some examples. During the eighteenth century, several astronomers including Titius and Bode realized that the semi-major axes of solar planets follow a regularity captured by the equation $a = 0.4 + 0.3 \times 2^n$, where a is the semi-major axis in astronomical units and n can be $-\infty$ (for Mercury) or an integer. This "law" produced empirically adequate results for the positions of all known planets at the time, and furthermore, resulted in novel predictions of the locations of Uranus and Ceres prior to their discovery.

How should one explain the novel predictive success of the Titius–Bode law? First of all, an extrinsic explanation might draw on the process of discovery of the formula and claim that the law was arrived at through a long numerical exercise and many trials and errors (probably a historically inaccurate claim!), and thus that its success is not surprising. If one should want to attempt an intrinsic explanation, on the other hand, one would have to compare the content of the theory and its predictions with the truth. Now, by contemporary lights, we do not know of any law or regularity that governs the locations of the planets, or of a coherent way to assign probability measures to initial conditions for that matter. As such, according to contemporary physics, initial conditions of the planets must be treated simply as matters of brute historical fact, and any pattern among them is likely a mere coincidence. If this is correct, then "the true theory" would not contain any formulaic procedure for deriving the same predictions as the Titius–Bode law. One would simply have to show that the specific numbers obtained through observations of planets' locations (excluding that of Neptune) happen to line up with the first few terms of the Titius–Bode sequence. There can be no deeper explanation that penetrates the level of unobservable reality. Therefore, the best explanation for the success of the Titius–Bode law is probably an intrinsic but unrobust explanation.[4]

By contrast, consider Niels Bohr's atomic theory. This theory was robustly successful: it returned correct predictions for all single-electron systems (neutral hydrogen atom, ionized hydrogen molecule, once ionized helium, twice ionized lithium, and the like) and in a wide range of conditions including in different excitations (initial states) as well as in the presence of external electric and magnetic fields (Stark and Zeeman effects, respectively). Bohr's theory broke down, however, as soon as spin came into play, which included phenomena related to helium and heavier elements, the hydrogen molecule, and the anomalous Zeeman effect. Nevertheless, as said above, the success of Bohr's theory was robust enough with respect to a fairly wide range of conditions, so much so that it provided the foundation for its extremely successful successor, non-relativistic quantum mechanics. As such, the success of Bohr's theory likely calls for a robust explanation. I will present one such explanation in §4.5 below.

[4]See (Lyons [2006], pp. 551–2) for further discussion of the Titius–Bode law.

How does one generally provide a robust explanation of the success of a given theory? First off, suppose we knew the true "God's eye" account of the nature of the physical system of interest, that is a (the?) theory all of whose terms correspond to objects of reality one-to-one—call it "the true theory"—and suppose we are attempting to explain the success of a given false theory through our knowledge of the true theory. A robust explanation of this success would begin by specifying a concrete range of conditions, typically in the form of constraints and resolution limits, and proceed to show that there is a formulaic relationship between the laws of the false theory and those of the true theory that remains stable to variations within said conditions.

In the absence of access to "the true theory", one may still provide a working explanation of the robustness of a false theory's success by comparing the theory to a much better successor. In such an exercise, one takes the successor theory to be a surrogate for the true theory (that is, one imagines a world in which the successor theory is literally true) and goes through the same procedure as above. Next, one examines whether the robustness of the successor theory can in turn be explained by appealing to its successors, and so on. That is, one would repeat the steps, this time taking what was surrogate truth as a false but robustly successful theory and one of its superior successors as the surrogate truth. Since robustness is transitive, if this cascade of robust relationships can be demonstrated, it would strongly suggest the existence of a robust relationship between the original false theory and elements of unobservable reality. To be sure, this cannot be established with complete certainty unless one arrives at "the Truth", but it can suffice to convince us of a robust relationship in the meantime.

My overall aim in this chapter is to argue that, contrary to the antirealist, a robust explanation is typically available for the success of dynamical theories, but one that, contrary to the realist, does not appeal to referentially transparent notions such as truthlikeness, representation, and structural similarity. The possibility of such a compromised explanation has been either ignored or outright denied by both realists and antirealists alike, which brings me to the next section.

4.3 The Fundamental Dichotomy and the Middle Path

4.3.1 Referential transparency

Metaphorically speaking, "referential transparency" is typically characterized by the idea of "reflecting" or "mirroring" reality. The reflection may be of entities, properties, structures, or what have you, but the point is that the way things truly are is transparently carried over to the way things are in the theory. Less metaphorically, I am speaking of the assumption that the hidden ontology or structure of the world can be read off the essential posits of at least some formulation of the theory.

> Referential transparency: There is some formulation of the theory such that the ontology or structure of (unobservable) reality can be read off the essential posits of the theory in that formulation.

A few clarificatory notes on referential transparency are in order. First of all, referential transparency does not imply that a unique ontology must be read off the theory. Consider the different interpretations of quantum mechanics such as Everettian and Bohmian QM.[5] Realist adherents of all these different interpretations disagree on which formulation is correct, but they do not disagree on whether the correct formulation of the theory—whatever it may be—is referentially transparent, which is to say the ontology or structure of the correct interpretation of QM reflects the ontology or structure of the world.

Secondly, even considering a single formulation of a given theory, different adherents of that formulation can disagree on the true ontology of reality if they disagree on what the essential posits of the theory are. The idea of essential posits (originating in Kitcher and Psillos) is that every theory comes with a set of core assumptions ("essential posits") to which the theory owes its empirical success, plus potentially other extra baggage ("idle posits") that play no role in the success of the theory. For instance, Kitcher ([1993], p. 143) argued that the existence of luminiferous ether was inessential to the empirical success of nineteenth-century electromagnetism, and therefore that a realist commitment to those theories does not commit one to the existence of the ether. Kitcher still assumes that the

[5] I am indebted to an anonymous reviewer for this example.

theory is referentially transparent: that whatever part of the theory is responsible for its empirical success does enable us to directly read off the (unobservable) ontology of the world; the disagreement is simply on which part is responsible for empirical success.

Truthlikeness, which is closely associated with all manner of scientific realism (including structural realism), represents the most common and arguably strongest form of referential transparency. As Chakravartty explains: "Epistemologically, realism is committed to the idea that theoretical claims (interpreted literally as describing a mind-independent reality) constitute knowledge of the world" ([2017], §1.2). Boyd similarly argues that approximate truth involves interpreting scientific theories "realistically" ([1983], p. 45) and "at face value" ([2002], §1).[6] The thesis of semantic realism thus implies a particularly strong form of referential transparency: it states that there is some formulation of the theory – namely the literal formulation – from which the ontology or structure of (unobservable) reality can be read off.

4.3.2 The Fundamental Dichotomy

4.3.2.1 Statement of the Dichotomy

It has been (tacitly or explicitly) assumed throughout the literature that there is only one way to provide a robust explanation for the success of a theory: by appealing to some referentially transparent notion such as truthlikeness, representation, or structural similarity between the theory and the unobservable world. Consequently, the literature is characterized by a widespread presupposition that if realism in the sense of referential transparency turns out to be unacceptable, then one has no choice but to default to unrobust explanations and bracket any discussion of how the intrinsic features of the theory compare to features of unobservable reality. I shall call this the Fundamental Dichotomy:

> Fundamental Dichotomy: Referentially transparent explanations are the only robust explanations of success. Any explanation of a theory's success must either appeal to referentially transparent relations, or to no robust world–theory relation at all.

[6] See also (van Fraassen [1980], §1.2); (Psillos [1999], p. xix); (Liston [2016], §5) on the role of literal interpretations in scientific realism.

Interestingly, as said above, the Fundamental Dichotomy is taken for granted by both realists and antirealists a claim I will defend in the following.

4.3.2.2 The Fundamental Dichotomy: the case of realists

What Musgrave ([1988]) refers to as "the Ultimate Argument" for scientific realism states that unless a predictively successful theory is approximately true and its theoretical terms referentially successful, the predictive success of the theory would have to be seen as an implausible sort of "miracle" or "cosmic coincidence". Since a reasonable explanation must avoid such cosmic coincidences, the argument goes, the best explanation for the predictive success of the theory is that it is approximately true and referentially successful.

The most commonly quoted passage in this context is from Putnam's *Mathematics, Matter and Method*, where he characterizes truthlikeness as "the only philosophy that doesn't make the success of science a miracle' (Putnam [1975], p. 73). The following, less commonly quoted passage from *Meaning and the Moral Sciences* is more elaborate:

> And the typical realist argument against idealism is that it makes the success of science a miracle [. . .] And the modern positivist has to leave it without explanation (the realist charges) that "electron calculi" and "space-time calculi" and "DNA-calculi" correctly predict observable phenomena if, in reality, there are no electrons, no curved space-time, and no DNA-molecules. If there are such things, then a natural explanation of the success of these theories is that they are partially true accounts of how they behave [. . .] But if these objects don't really exist at all, then it is a miracle that a theory [. . .] which speaks of curved space-time successfully predicts phenomena. (Putnam [1978], pp. 18–9)

One can also find very similar passages in earlier writings by Smart (see ([1963], p. 39); ([1968], p. 150)), who in addition to "miralces", often speaks of "cosmic coincidences" and "innumerable lucky accidents' as alternatives to realism.

These passages by Smart and Putnam are puzzling to say the least. In uncharacteristic fashion for a piece of analytic philosophy, no explicit argument is offered. One rather finds a series of rhetorical questions sandwiched by mere assertions. And to make matters worse, the assertions themselves involve several mysterious, undefined terms such as "inexplicable", "without explanation", "coincidence", and "miracle". One is left wondering: what is bad about coincidences and miracles that disqualifies them as explanations? And what is the connection between all these terms that licenses moving from one to another?

Let us start with "inexplicable". Since the argument is typically couched in a framework of inference to the best explanation, these assertions are typically interpreted as saying that the best explanation of scientific success is approximate truth. Under such an interpretation, when Smart and Putnam claim that barring realism, scientific success would be left "inexplicable" or "without explanation", one must read them as meaning to say "without plausible alternative explanation". After all, a coincidence (or a "miracle" for that matter) is an explanation, too. In fact, in some cases coincidence is the best explanation (say if I run into an old friend while thinking about them). However, the idea goes, in the case of scientific success, mere coincidence is not a good explanation. Under this reading, all that matters about terms such as "coincidence" and "miracle" is that they are extremely improbable events, which I take it is why they are bad explanations.

The above reading would remove part of the puzzle about these passages, but one is still left wondering: what justifies the swift move between "without (plausible alternative) explanation" on the one hand, and "coincidence" or "miracle" on the other? Are Smart and Putnam equating "coincidence" with "inexplicable event"?[7] Clearly, not all implausible alternative explanations of success must draw on coincidences. Here is one implausible explanation of the success of science that does not draw on coincidences in the sense of extremely unlikely events: God had arranged the world in a pre-established harmony to make sure that the false theories appealing to human psychology are also predictively successful, without the theories having any intrinsic relationship to unobservable reality. And of course less outlandish examples are found in the methodological explanations advocated by antirealists (see below). This becomes even more puzzling once one realizes that Putnam himself was aware of methodological explanations of success, and spent considerable effort to critically evaluate them in other contexts (Putnam [1981], p. 195; [1963a/1975], p. 280; [1963b/1975], p. 295).

The fact that Putnam makes no reference to methodological explanations in the context of No Miracles suggests that he does not see methodological explanations as rivals to the type of explanation he is looking for in the passages above. Putnam must be demanding

[7]Van Fraassen assumes this reading: "it is illegitimate to equate being a lucky accident, or a coincidence, with having no explanation" ([1980], p. 25).

"explanation" in a specific sense that cannot be substituted by any amount of methodological explanation, for it does not compete with it. This special sort of explanation, whatever it may be, must be "non-coincidental". This would explain Smart and Putnam's apparent jump from "inexplicable" to "coincidental".

This makes the passages above much more plausible, but at one cost: one must assume that approximate truth is the only non-coincidental explanation of success. Now, although the aim of this chapter is precisely to argue against the latter assumption, one can see how Smart and Putnam would have been led to it. They take the essential content of a scientific theory to be a static description of entities, properties, and/or structures. And the most obvious—arguably the only—robust relation a static description can have with the thing it describes is for the description to be (approximately) true of the thing described, and/or for the terms of the description to somehow mirror the thing described.

The assumption that referentially transparent explanations are the only robust explanations of success has subsequently become far more explicit in the realist literature. Niiniluoto, who has perhaps made the most significant attempts at rigorously defining truthlikeness[8], puts it most unequivocally:

> [A]n explanation of the ability of a scientific theory to yield successful predictions ... has to refer to some permanent property of the theory, which describes its relation to the world. Truthlikeness is the best—even the only—property I know that could serve this function. (Niiniluoto [1999/2002], p. 198)

In my terminology, the requirement that the explanation of success "refer to some [. . .] property of the theory" is the requirement that the explanation be intrinsic. More specifically, the condition that the explanation of success refer to a "permanent" property which "describes [the theory's] relation to the world" entails a robust explanation. Under this reading, the last sentence of the quote above effectively asserts a version of the Fundamental Dichotomy.

The Dichotomy is also implied when realists swiftly move from some theory–world relationship to truthlikeness specifically. For instance, Leplin implies that the only way to "capture" some features of natural processes is to "reveal some significant truth" about them ([1997], p. 104). The situation is similar with Worrall:

[8]See (Niiniluoto [1987]; [1998]; [2003]).

> How, it seems natural to ask, could Fresnel's theory correctly make a prediction [of the bright spot at the centre of the shadow of a disk] that is so at odds with what "background knowledge" would lead us to expect, unless it had somehow or other latched on to the way that light really is? The theory, it seems natural to conclude, must be at least approximately correct if it can get such a striking prediction right. (Worrall [2007], p. 126)

There is a very quick transition in this passage between the idea that the theory "latches on to" the inner workings of light and the idea that the theory makes correct descriptive assertions about light (in Worrall's case, about the "structure" of light). This creates the impression that except through truthlikeness, there is no way to flesh out this relation of "latching on to". Once again, if we let phrases such as "latch on to" and "capture" be metaphors for robust theory–world relationships, Worrall and Leplin's statements amount to saying that the only way for the theory to have a robust relationship with the world is truthlikeness, which is a version of the Fundamental Dichotomy.

Incidentally, this highlights the stark contrast between my middle path proposal and structural realism, which was also introduced as some sort of "third way" or "the best of both worlds": the structuralist insists on keeping some (attenuated) notion of truthlikeness and referential transparency. In essence, the structural realist takes the lesson of the No Miracles argument to be truthlikeness, and that of Pessimistic Induction to be a denouncing of entities. However, I argue that we should take the lesson of No Miracles to be the necessity of a robust theory–world relation (of which truthlikeness is a special case), and the lesson of Pessimistic Induction to be a denouncement of all referential transparency, whether it pertains to entities, properties, structures, or states.

One can infer the tacit assumption of the Fundamental Dichotomy from many other realist texts, but the point has been made.[9]

In the next section, I will argue that the presupposition of the Fundamental Dichotomy has been as prevalent among antirealists as it is among realists.

[9]See (Park [2001], Ch. 3), (Niiniluoto [2018], p. 163), (Alai [2014]). Recently, Williams has proposed what he calls effective realism, which cautions against simplistic or literal readings of theories. Nevertheless, Williams continues to hold on to referential transparency for a select group of entities picked out through renormalization techniques (see (Williams [2019], p. 221)).

4.3.2.3 The Fundamental Dichotomy: the case of antirealists

The methodological explanation of scientific success, pioneered by van Fraassen ([1980]), Wray ([2007]; [2010]), and Laudan ([1984]), goes something like the following. It is a brute fact that occurrences in our world exhibit regular, repeatable patterns. Some of these regularities are bound to percolate up to the level of observables and generate patterns of observable regularities. A theory that captures all these observable regularities is called empirically adequate. Since these are brute facts, the only question left to address is: how do we come up with such successful theories? The answer according to the methodological view is that our successful theories are successful because they have gone through a brutal process of modification with very high selection pressure—"a jungle red in tooth and claw" (van Fraassen [1980], p. 40). New modifications of our ideas are being presented all the time and the ones that ever so slightly deviate from observable results are discarded. It is no surprise, then, that the ones that survive turn out to be pretty empirically adequate.

The mark of the Fundamental Dichotomy can be found in the classic passages in which van Fraassen introduced the idea of an "evolutionary" explanation of success:

> I can best make the point by contrasting two accounts of the mouse who runs from its enemy, the cat. [The first is:] the mouse perceives that the cat is its enemy, hence the mouse runs. What is postulated here is the "adequacy" of the mouse's thought to the order of nature: the relation of enmity is correctly reflected in his mind. But the Darwinist says: Do not ask why the mouse runs from its enemy. Species which did not cope with their natural enemies no longer exist. That is why there are only ones who do. (van Fraassen [1980], p. 39)

As the passage above makes clear, van Fraassen envisions two possibilities: either something in the mouse's brain accurately reflects something in the world, or else there is nothing to be said about what is on the mouse's mind. The former option represents a referentially transparent explanation, with the word "reflect" signifying the idea of mirroring unobservable reality in a transparent manner. The quietist option, on the other hand, represents giving up on trying to find any robust relationship between the mouse's mental goings on and facts of reality. This false dichotomy serves to hide the third, middle path option: that something in the mouse's brain robustly tracks something in reality within the typical contexts in which the mouse finds itself, without that "something" having to "reflect" anything in reality.

Laudan also reveals his commitment to the Fundamental Dichotomy in the course of discussing his own methodological explanation of success:

> When we ask why scientific theories work so well, we might be asking [. . .] to be told what semantic features theories possess in virtue of which they have such an impressive range of true consequences. Alternatively, when we ask why science is successful, we might be asking an epistemic and methodological question about the selection procedures which scientists use for picking out theories with such impressive credentials. If, as I suspect, it is generally the latter which we are driving at, then the realist response becomes even less availing than it already appears to be [. . .] (Laudan [1984], pp. 91–2)

Laudan seems to be presenting us with the following choice: either a semantically defined property of the theory explains its success (which will lead to the truthlikeness thesis), or the explanation must draw on no world–theory relationship at all, but only on a theory–theory relationship (namely having been arrived at through stringent testing of a former hypothesis). Once again, we are told that if truthlikeness doesn't work, then the world–theory relation must remain mysterious.[10]

Stanford ([2000]) does not directly appeal to methodological explanations, but rather to the notion of predictive similarity.[11] Nevertheless, Stanford also relies on presupposing the Fundamental Dichotomy. He writes:

> [There is] no clear reason to expect a successful novel prediction to be much more likely (if more likely at all) to come from an approximately or partially true theory than to come from a false theory that has nonetheless managed to make predictions sufficiently similar to the theoretical truth of the matter [. . .] (Stanford [2000], pp. 281–2)
> [T]he proposal offered here does not appeal to a relationship between a theory and the world at all; instead it appeals to a relationship of predictive similarity between two theories. (p. 276)

Granted that there is no reason to assume successful novel predictions to come from approximately or partially true theories, why should the alternative on offer give up on all world–theory relationships altogether? Once again, it is the Fundamental Dichotomy at work.

Other examples can be proliferated *ad nauseum*[12], but the above should suffice to show that antirealists have been committed to the Fundamental Dichotomy.[13]

[10] See (Wright [2002]; [2014]) for a slightly different methodological explanation of success.
[11] See (Psillos [2001]), (Park [2003]), (Musgrave [2007]) for discussion.
[12] See (Lyons [2003]); (Hoyningen-Heune [2011]); (Dawid [2018]).
[13] Arthur Fine has famously suggested that we break free from both realism and antirealism, but as far as

I have argued that both realists and antirealists widely adhere to the Fundamental Dichotomy. Next let us see why the Dichotomy must be rejected.

4.3.3 Problems with the Dichotomy and the prospect of a middle path

There are many problems with the assumption of the Fundamental Dichotomy. In particular, there are three reasons for rejecting it.

For one thing, existing philosophical theories which rely on the two horns of the Dichotomy are notoriously inadequate in meeting the demands of both No Miracles and Pessimistic Induction arguments: many robust explanations that rely on referentially transparent notions (for instance entity realism) are too demanding to pass the test of history; others involve technical terms that are so watered down and fragmented (for instance "structure" in structural realism) that it is unclear what substantive commitments are entailed by the view. On the other hand, unrobust explanations allow that the matching regularities of observable and theoretical terms be taken as "merely a brute fact" (van Fraassen [1980], p. 24; see also (Boyce [2018])). This dodges the historical counterexamples, but leaves one deeply unsatisfied on No Miracles grounds.

Secondly, the Dichotomy does not exhaust the logical space. For while a referentially transparent theory–world connection is sufficient for removing the appearance of miraculous success, it is not necessary. A robust formulaic connection, on the other hand, is arguably both necessary and sufficient for the same purpose. On the flip side, while denouncing all robust theory–world relations is sufficient for avoiding historical counterexamples, it is not necessary. A referentially opaque relation, on the other hand, is arguably both necessary and sufficient for the same. Consequently, insofar as "realism" requires commitment to referential transparency and "antirealism" requires bracketing of all robust theory–world relations, realism and antirealism do not exhaust the space of logical possibilities.

Thirdly and most importantly, I maintain that the Dichotomy can be shown to be demonstrably false through case studies. One can present examples from the history and practice

explaining the success of scientific theories goes, Fine resorts to a form of "quasi-realism" or "surrealism" (Leplin [1987]), according to which the world behaves as if the theory is true (Fine [1986], p. 160), which is an unrobust explanation of success.

of physics where a robust but referentially opaque relationship can be demonstrated between the theory and a much better successor, and between the successor and its successor, and so on, which given transitivity strongly indicates a robust-yet-opaque relationship to "the true theory". Such case studies thus serve to motivate a middle path between realism and antirealism from both directions: they show both that a theory's success is owed to intricate connections to unobservable reality and that these connections fall short of referential transparency.

Consider the example of Bohr's atomic theory. Bohr ([1913]) posited that an atom is analogous to a mini-solar system, the nucleus being the "sun" and the electrons orbiting it being like "planets", but with one important restriction: of all the orbits that are classically permitted by Coulomb's inverse square law, the electron may occupy only those that satisfy the constraint

$$E_n = -\frac{R}{n^2}, \tag{13}$$

where E is the binding energy of the electron, R is the Rydberg constant, and n is an integer known as the principal quantum number, which labels the discrete energy levels. Bohr called these discrete permissible states "stationary states". Bohr further posited that the electron does not emit radiation unless it "jumps" from one stationary state to another, in which case the difference of energy will be released as radiation of frequency

$$\nu_{nm} = \frac{1}{h}(E_n - E_m). \tag{14}$$

Using these state assignment rules, Bohr was able not only to explain the line spectra of hydrogen, but also to make impressive, novel predictions about the spectral lines of once ionized helium, twice ionized lithium, and other single-electron ions. These successes reportedly prompted Einstein to say "the theory of Bohr must then be right!" (See (Pais [1991], p. 154).)

The empirical data to which Bohr responded was spectral lines—little bands of light emitted from stimulated atoms and molecules. Since these lines are characterized by frequency and intensity, and since frequencies and intensities can be inferred from suitable

state assignments, Bohr's theory owed its success solely to its state assignments.[14] In particular, the ontological assumptions about the mini-solar system and the Kepler orbits play no more than a heuristic role in the theory (see (Norton [2000], pp. 86–7)).[15]

Importantly, the source of Bohr's success is not inferred from comparison with successor theories or otherwise through the benefit of hindsight. Indeed, as early as the 1920s, the idea of dispensing with classical orbits and focusing on state assignments was gaining popularity among physicists including Bohr himself. Van Vleck discussed whether this "bold proposal" would "invalidate" the successes of Bohr's theory and concluded in the negative (Van Vleck [1926], pp. 108–9). Similar arguments regarding the dispensability of ontology and essentiality of state assignments for Bohr's success were made by Born ([1925], pp. 113–4), Kramers ([1924], p. 311), and Heisenberg (letter to Kronig 6/5/1925).

Thus, one can directly demonstrate that the success of Bohr's theory was not due to the referential success of its mini-solar system ontology. Hindsight, while unnecessary, certainly confirms this: if our current theories are on the right track, then the ontological descriptions of Bohr's theory were completely unreferential: there are no such things as little grains of charged dust swirling around the nucleus in classical multi-periodic orbits much like a mini-planetary system. Consequently, there is no interesting sense, and certainly no transparent sense, in which Bohr's theoretical entities, properties, or structures refer to or represent unobservable reality. This rules out one horn of the Dichotomy.

At the same time, one can also show that Bohr's stationary states have a robust (albeit referentially opaque) connection to the Schrödinger state assignments, and the latter to quantum field theory, and so on (see §4.5 for proofs). The success of Bohr's model is therefore not a "coincidence" or simply due to the meticulous process of theory selection: there is an intrinsic reason why Bohr's state assignments work, namely that they are connected to the true state assignments in the right way. As I explain below, this connection is called "tracking".

As such, my explanation of the success of Bohr's theory goes beyond traditional antirealist

[14]Spectral lines were also found to have polarization and phase. These features can easily be derived from Bohr's state assignments as well.

[15]For further discussion, see (Vickers [2012]; [2018]).

explanations of success and falls short of traditional realist explanations.[16]

We are now ready to revisit the No Miracles argument in light of the above considerations.

4.4 Taking Stock: a Modified No Miracles Argument

Traditionally, the No Miracles argument is characterized along the following lines (from (Niiniluoto [2018], p. 159)):

> P1: Many theories in science are empirically and pragmatic [sic] successful.
> P2: The truth or truthlikeness of scientific theories is the best explanation of their empirical and pragmatic success.
> C: Hence, conclude that such successful theories are truthlike.[17]

I find this formulation unhelpful, in part because the way P2 is phrased, it says nothing about the implicit demand that the explanation be intrinsic and robust, and it masks the tacit assumption of the Fundamental Dichotomy. To remedy these defects, I propose that the original No Miracles argument be formulated as follows:

> Q1: Mature scientific theories are robustly successful.
> Q2: Unless the theory were referentially transparent, the success of the theory would be a confluence of countless coincidences (a "cosmic coincidence").
> Q3: Cosmic coincidences are extremely improbable events ("miracles").
> C: In all probability, mature scientific theories are referentially transparent.

To be clear, in this reading "coincidence" is not defined as an unexplained or extremely unlikely event, but rather as the line-up of two events that have no intrinsic or robust relationship to each other. A "cosmic coincidence" is simply defined as a confluence of numerous coincidences. A "miracle" is then defined as an extremely unlikely event. With that in mind, Q2 follows from the fact that our theories are robustly successful (hence the "confluence") plus the assumption of Fundamental Dichotomy. Q3, on the other hand, is proposed as an empirical claim (not a matter of definition).

[16]The false character of the Fundamental Dichotomy has recently been recognized by Kyle Stanford in a 2017 talk in honour of Philip Kitcher's seventieth birthday in which he introduces the phrase "middle path" (Stanford [unpublished]; see also ([2018])).

[17]See also (McCain and Kampourakis [2019], pp. 246–7) for a similar formulation.

As the textual evidence above indicates, all efforts on the antirealist side have been focused on refuting Q3: antirealists have argued that the cosmic coincidence of robust success is not improbable at all, either because history presents us with many clear cases of it or because science follows a rigorous methodology that is designed precisely to produce such cosmic coincidences. As I have been arguing, however, the bigger problem with the traditional No Miracles argument is Q2 (not Q3), because it presupposes a false dichotomy. In light of this, I propose the following, Modified No Miracles argument:

> R1: Mature scientific theories are robustly successful.
> R2: Unless the theory were robustly related to unobservable reality, the success of the theory would be a confluence of countless coincidences (a "cosmic coincidence").
> R3: Cosmic coincidences are extremely improbable events ("miracles").
> C: In all probability, mature scientific theories are robustly related to unobservable reality.

In this argument, R2 follows directly from the fact of robust success (expressed in R1) plus the definition of "coincidence" (which is defined as lack of robust relationship). Thus, unlike the deeply problematic Q2, R2 is indisputable (indeed it can be removed from the argument as it adds no new information). R3, on the other hand, maintains the same status as Q3: it is put forth an empirical claim. Therefore, as long as one accepts that science is robustly successful, the entire force of the Modified No Miracles argument stands and falls with R3.

Perhaps there are reasons to doubt R3. But my goal here is not to question that premise. Rather, I wish to argue that one can comfortably take Modified No Miracles as a forceful argument and respond to its force without positing problematically transparent notions of reference. The task is therefore to find an explanation of success that makes sense of the idea that the theory "gets something right" or "latches on to" or "captures" something about the inner workings of the system, and yet avoids commitment to referential transparency. This is the topic of the next section.

4.5 Tracking: Robust but Referentially Opaque

4.5.1 The basic features of the tracking relationship

4.5.1.1 Tracking defined

I propose that the dynamical systems framework allows us to solve the problem of success by providing a robust world–theory relation that explains success but without referential transparency (even of the retrospective kind). This relationship is called tracking. Given two dynamical systems $\mathrm{DS}_1 = (S, D_\delta)$ and $\mathrm{DS}_2 = (S', D'_\delta)$, where S and S' are sets of states and D_δ and D'_δ transition operators, we have:

> Tracking: DS_1 tracks DS_2 if and only if there is a mapping from S' to S (inducing equivalence classes on S') such that members of each equivalence class defined on S' transition in a manner that is indistinguishable according to the same state assignment rules at least within a range of contexts.[18]

More formally: if we denote the mapping by $m : S \to S'$, we say that DS_2 tracks DS_1 if and only if the mapping and the transition operators "cooperate" (or "commute") in the sense that

$$\forall s \in S, m(D_\delta(s)) \approx D'_\delta(m(s)). \tag{15}$$

In other words, DS_2 tracks DS_1 if and only if the result of evolving an S-state and then mapping it to S' is the same as mapping to S' first and then evolving the S'-state.[19]

Tracking means that these equivalence classes are dynamically preserved within the appropriate range of contexts, which means that all states that are in the same equivalence class as each other will transition to other states that are also in the same equivalence class as each other. In other words, if two "lower" states are mapped to the same "higher" state,

[18] The term "tracking" appears in a few other places, including the title of Psillos's book: *Scientific Realism: How Science Tracks Truth* and "tracking theories" in epistemology associated with Dretske, Nozick, and others (see: (Dretske [1970], [1971], [1980]); (Nozick [1981]); (Roush [2005])). The similarity in terminology is purely linguistic: Psillos is a straightforward entity realist who believes in the strongest form of referential transparency, and the "tracking theories" in epistemology concern tracking the truth value of propositions, not dynamical states.

[19] Variants of tracking have been defined by Wallace ([2012]), Rosaler ([2015a]), Yoshimi ([2012]), and Guinti ([2006]).

then the "lower" dynamics takes the two lower states to other states that are also mapped to the same higher state.

I will showcase a few examples of tracking below, but first let us briefly explore some of the unique features of tracking that set it apart from other explanations of success.

4.5.1.2 Tracking is referentially opaque

One might be tempted to say that tracking is a form of referential success. After all, does tracking not imply that the higher-level states refer to those lower-level states that transition indistinguishably? In other words, are the theoretical terms not picking out certain natural groupings of the underlying states that evolve similarly?

The answer is no. Since tracking depends on the interaction between the underlying transitions and the mapping between the two sets of states, it cannot be reduced to any static or atomistic referential relationship. There is nothing natural or inherent about the groupings themselves: the equivalence classes do not share any property that could be described purely at the level of truth and without reference to the higher-level theory. To see this, first note that successful state assignments are not unique. Consider a deterministic system with four true states $S_1, ..., S_4$ and let the transition rules be $S_1 \rightarrow S_3$ and $S_2 \rightarrow S_4$, and suppose the characteristic magnitudes of the two transitions are roughly equal. Now consider the following three ways of mapping ("lumping") the underlying states into equivalence classes of theoretical states:

1) Assign $\{S_1, S_2\} \leftrightarrow S_T$, $\{S_3, S_4\} \leftrightarrow S'_T$
2) Assign $\{S_1, S_2\} \leftrightarrow S_T$, $\{S_3\} \leftrightarrow S'_T$, $\{S_4\} \leftrightarrow S''_T$
3) Assign $\{S_1\} \leftrightarrow S_T$, $\{S_2\} \leftrightarrow S'_T$, $\{S_3\} \leftrightarrow S''_T$, $\{S_4\} \leftrightarrow S'''_T$

It is easy to verify that the coarse-grainings in (1) and (3) result in state assignments that track the underlying dynamics, whereas the ones in (2) fail to do so. Now compare the assignment $\{S_1, S_2\} \leftrightarrow S_T$ in (1) to the same assignment in (2). Both assignments lump the same underlying states into the same equivalence class, yet one is successful and the other is not. Clearly, the success of (1) is not due to the fact that S_T "picks out a real property" of the system, or else the same assignment would not fail in (2). The proponent of referential success might try to attribute this to the fact that (2) fails to lump S_3 and S_4 together.

Perhaps by failing to put the latter two states in the same equivalence class, (2) fails to represent an important property that they share and thus fails to predict. In other words, perhaps (2) is being too fine-grained for its level. But this cannot be right, because the same fine-grained assignment works perfectly well in (3). In short, the success and failure of state assignments cannot be decided in isolation. It is the entire set of state assignments that either tracks or fails to track as a whole.

As we have seen, tracking is irreducible to referential success because it is the result of a confluence of many factors including the way the underlying states are mapped to theoretical states and the underlying dynamics, not to mention the constraints on the system and our experimental resolution limits.

4.5.1.3 Tracking is stronger than empirical adequacy

I have argued that tracking is weaker than referential success. For it to be a genuine third option, tracking must also be stronger than mere empirical adequacy. This is indeed the case: tracking is a relation involving theoretical states, and the latter are typically unobservable. To use the Bohr example again, what one directly observes (measures) are little bands of light called spectral lines. According to Bohr's theory, the frequency of each line is given by the difference of two energy states: $\nu_{nm} = \frac{1}{\hbar}(E_n - E_m)$. Thus, one always observes ν_{nm}, but never E_n or E_m itself. Since the theoretical states are unobservable, tracking expresses more than a relationship among observables and is therefore stronger than mere empirical adequacy.

Indeed, tracking would be stronger than empirical adequacy even if the theoretical states were directly observable. Consider a theory of macroscopic objects in which the state of the system simply consists of the position of the object (an observable). Is tracking distinct from empirical adequacy in that case? The easy answer would be yes, because in addition to the theoretical states, the relation described in a tracking claim has the true states as one of its relata, and the true states are presumably unknown to us.

Granted their conceptual possibility, what can we say about the actual existence of theories that are empirically adequate but do not track? The Titius–Bode law discussed in

§2 above provides an example: it is empirically adequate insofar as it predicts the locations of solar planets up to Neptune, but one that likely tracks no underlying reality insofar as it imposes law-like restrictions on what we now recognize as brute initial conditions with no deeper dynamics to track.

What if there is a deeper dynamics to track and the theory does posit a state space? Could such a theory be successful without its states tracking the true states of the system? The answer is yes. Consider the case of blackbody radiation. In the latter half of the nineteenth century, the quest was on to explain the spectrum of radiation from perfectly absorbing materials known as "blackbody objects" (typically a dark cavity with a hole) when in thermal equilibrium with the environment. More specifically, the focus was on experimentally derived relations that expressed the power irradiated as a function of temperature and frequency of radiation. The phenomenologically accurate formula for blackbody radiation was later discovered by Planck and is known as the "Planck distribution law":

$$P(\nu) \propto \frac{\nu^3}{e^{\alpha\nu} - 1} \tag{16}$$

where ν is the frequency of radiation, P is the power radiated per unit solid angle per unit area per unit frequency, and α is a temperature-dependent constant. However, a few years before Planck introduced his complete formula, Wien as well as Planck himself derived the following formula for blackbody radiation from classical thermodynamic considerations[20]:

$$P(\nu) \propto \frac{\nu^3}{e^{\alpha\nu}} \tag{17}$$

It is easy to see that for sufficiently large frequencies ($\alpha\nu \gg 1$), the -1 term in the denominator can be ignored and Planck's distribution law reduces to Wien's formula. As such, Wien's formula is empirically adequate at predicting radiated power for higher frequencies of radiation (in agreement with experimental results by Paschen and others at the time), although it breaks down in the low-frequency region.

However, Wien's theory does not track the true underlying dynamics of blackbody radiation even in the region in which it is empirically adequate (higher frequencies). The reason

[20]For a brief account of how Wien and Planck derived their respective formulae, see (Mehra and Rechenberg [1982], p. 24); (Crepeau [2009]).

is that the higher frequencies correspond to electron transitions near the nucleus, which is the proper realm of quantum mechanics. Wien, on the other hand, assumed a classical state space in accordance with Maxwell's molecular theory.[21] Taking quantum mechanics as the surrogate truth, one can see that the continuous state space of classical physics simply fails to track the discrete state space of the electron in this region.[22] Contrast this to a hypothetical theory QM* which prescribes the same state assignments as quantum mechanics for transitions near the nucleus but diverges from the latter for higher orbits. Such a theory would be as empirically adequate as the Wien formula, but unlike the latter, it would also track the true dynamics of the system in the high-frequency regime.

To summarize, I have argued that tracking is stronger than empirical adequacy both on conceptual grounds and on the basis of concrete examples.

This concludes my general remarks on the properties of tracking. In the following, I will explore several concrete examples of tracking.

4.5.2 Fulfilling the promise: some examples of tracking

In this subsection, I will explore several examples where a theory tracks a successor. The examples show the simultaneous rigour and flexibility of the notion of tracking, and exemplify my assertions above regarding the cascade of robust relationships.[23]

One can provide a general format for tracking explanations as follows:

> Steps to providing a tracking explanation of the success of a (dynamical) theory:
> 1- Specify constraints and resolution limits that delineate the context / range of robustness of successful theory.
> 2- Demonstrate a mapping between the state assignments of the (surrogate) true theory and those of the theory in question.

[21] Note that the phrase "Wien's approximation" can be misleading: the formula was theoretical, not approximate or phenomenological.

[22] I will not go into the proof of the latter statement, but the argument is roughly as follows: For the continuous state space of classical mechanics to track the discrete state space of quantum mechanics, certain constraints must be met (see (Rosaler [2015b]) for details). One of these constraints is that the fluctuations in the force field be negligible within the spread of the wave packets in the system: $\langle \frac{\partial V}{\partial x} \rangle \approx \frac{\partial V(\langle x \rangle)}{\partial \langle x \rangle}$. This condition fails near the nucleus where the Coulomb force varies dramatically from one orbit to another, which is the source of the higher frequencies.

[23] One example that I shall not explore due to limited space is that of classical mechanics tracking non-relativistic quantum mechanics. The main elements of the proof of this relation can be found in (Rosaler [2015b]). See also (Wallace [2012]).

3- Show that the mapping from step 2 is dynamically preserved by (aka "commutes with") the underlying "true" transition rules within the constraints and resolution limits specified in step 1.

Let us see how this works.

4.5.2.1 Old quantum theory tracks non-relativistic quantum mechanics

Let the nth Bohr energy state be denoted by S_n and let the quantum mechanical states be Hilbert states of the form $|\psi_n\rangle = |\phi_n\rangle\,|s\rangle$ (where $|\phi_n\rangle$ is the spatial part and $|s\rangle$ the spin part of the Hilbert state with principal quantum number n). Taking quantum mechanics as the surrogate truth, we have the following robust explanation for the success of Bohr's theory:

1.1. Constraints: The system contains no more than one electron.

1.2. Resolution limits: Spin-related splittings are ignored.

2. Mapping: Let $m(|\psi_n\rangle) = \langle\psi_n|\,H^*\,|\psi_n\rangle = E_n$, where H^* is a Hamiltonian operator from which all spin-related terms have been eliminated, and E_n is the energy contained in the spatial part of the state.

We must show that $S_n = E_n$, or that the mapping indeed takes us from Hilbert states to Bohr states. In the case of a central Coulomb force (which was the primary source of Bohr's success in terms of spectral lines and most other experiments of interest), the mapping can be demonstrated directly by calculating the states in both theories and showing that $S_n = E_n = \frac{\hbar R}{n^2}$, where R is the Rydberg constant and n the principal quantum number. More generally, one can convince oneself that the two theories lead to similar state assignments for systems with one degree of freedom and no spin, by noting that the quantum mechanical commutation relation $[q, p] = i\hbar$ can be obtained from the Bohr–Sommerfeld quantization condition $\oint p dq = n\hbar$ through Heisenberg's substitution recipes. However, it is not possible to provide a rigorous formal proof that the two sets of states coincide, and they may not for many potential functions other than that of the Coulomb force.

3. Robustness: We must show that within the relevant constraints and resolution limits (1.1 and 1.2 above) if $S_i = E_i$ then $S_f = E_f$ where the subscript i denotes initial states and f final states. The following considerations should reassure us that this is the case. First

of all, since there is only one electron in the system (condition 1.1), the energy of the bond between the single electron and the nucleus is the only relevant quantity in the experiments to which Bohr's theory answers. Thus, there are no extra degrees of freedom to which the energy of the electron can be transferred as a result of time evolution. Moreover, since spin-related interactions are ignored (condition 1.2), the Hamiltonian which generates the evolution operator will contain no spin-related terms, which means that states of the same spatial character but different spin would transition similarly within the accepted resolution. Therefore, the energy of the one spatial degree of freedom encodes all of the transition rules of interest in these contexts.

Secondly, since time evolution is linear in quantum mechanics, the transitions are fully deterministic, which implies that any two initial states with the same spatial energy content will transition to two final states with the same spatial energy content[24]. Since Bohr's states pick up the spatial energy content, this means that any two initial states that map to the same Bohr state will transition to two final states that also get mapped to the same Bohr state. This indicates that the mapping from surrogate truth to unobservable theory is robust within the constraints and resolution limits above. This completes our tracking explanation.

The above argument is admittedly rather qualitative and leaves out a direct comparison between the transition rules of the two theories. Normally one would show the robustness of the relation by appealing to the compactified versions of the two theories' transition rules (see examples below). However, this cannot be done in this particular case, because neither theory actually provides any way to compactify the transition rules of interest. Old quantum theory notoriously struggled with compactifying the intensity rules that provide the transition probabilities between pairs of Bohr's stationary states; and non-relativistic quantum mechanics cannot accommodate particle creation (or annihilation), which means that it has no transition rules for going from a state without a photon (the electron's initial state) to one in which a photon is emitted (the electron's final state). Therefore, one could not hope for a formal proof.

[24]By "spatial energy content" I mean the quantity $E_n = \langle\psi_n| H^* |\psi_n\rangle$, where H^* is a modified Hamiltonian without spin.

4.5.2.2 Free, spinless non-relativistic quantum mechanics tracks quantum field theory

Next let us consider the relationship between non-relativistic quantum mechanics and its successor quantum field theory. Let the "higher-level" states be spinless wavefunctions $\psi(x,t)$ whose transition rules are given by the free Schrödinger equation

$$\frac{\partial}{\partial t}\psi(x,t) = \frac{i}{2m}\nabla^2\psi(x,t). \tag{18}$$

Let the "lower-level" state assignments be given by states $|\phi\rangle = \int d^3k e^{iE(\vec{k})t}\,\alpha(\vec{k})\left|\vec{k}\right\rangle$ in the Fock space of a free Klein–Gordon field with the transition rules

$$\frac{\partial}{\partial t}|\phi\rangle = -\frac{i}{2}\int d^3y[\hat{\pi}^2 + (\nabla\hat{\xi})^2 + m^2\hat{\xi}^2]\,|\phi\rangle\,, \tag{19}$$

where $\hat{\xi}$ is the Hilbert operator associated with the Klein–Gordon field and $\hat{\pi}$ its conjugate momentum. Taking the latter theory as the surrogate truth, we have the following robust explanation of the success of the former theory:

1.1. Constraints: The spectrum of $|\phi\rangle$ quickly drops beyond a certain momentum $p_{\max}$, given by $\alpha(k \geq p_{\max}) \ll 1$.

1.2. Resolution limits: Fluctuations of order $\frac{p_{\max}}{m}$ are ignored; that is: $O(\frac{p_{\max}}{m}) \approx 0$.

2. Mapping: Let $m(|\phi\rangle) = e^{-i(m+E_0)t}\,\langle 0|\,\hat{\xi}(x)\,|\phi\rangle$, where E_0 is the vacuum energy.

Like before, we show the coincidence of the two sets of states simultaneously with the proof of robustness, which shows that $e^{-i(m+E_0)t}\,\langle 0|\,\hat{\xi}(x)\,|\phi\rangle$ satisfies the same equation of motion as $\psi(x,t)$.

3. Robustness: Using the above conditions and the transition rules (19) for the free Klein–Gordon field, we can show the following robustness result:

$$\frac{\partial}{\partial t}[e^{-i(m+E_0)t}\,\langle 0|\,\hat{\xi}(x)\,|\phi\rangle] \approx \frac{i}{2m}\nabla^2(e^{-i(m+E_0)t}\,\langle 0|\,\hat{\xi}(x)\,|\phi\rangle). \tag{20}$$

A proof of (20) is provided in (Rosaler [2015a], Appendix). Comparison of (18) and (20) completes our proof that a free Schrödinger particle tracks a free Klein–Gordon field within the conditions above. The case of an interacting (as opposed to free) quantum field theory is substantially more difficult and will not be discussed here.

Note: the foregoing is an illustration of the "cascade" of tracking relations discussed in §4.2 above: Bohr's atomic theory tracks non-relativistic quantum mechanics, but the latter theory itself tracks relativistic quantum field theory. Thus, since tracking is transitive, Bohr's atomic theory also tracks quantum field theory.

However, the argument above only deals with spinless states. We need a separate argument to show that non-relativistic spin states also track relativistic quantum mechanics.

4.5.2.3 Spin states track relativistic quantum mechanics

Let the "higher-level" states be two-spinors of the form

$$|s\rangle = \begin{pmatrix} |\psi_1\rangle \\ |\psi_2\rangle \end{pmatrix} \tag{21}$$

whose transition rules are given by the Pauli equation:

$$\frac{\partial}{\partial t}|s\rangle = -\frac{i}{\hbar}\big[\frac{1}{2m}(\vec{\sigma}.(\hat{p} - q\hat{A}))^2 + q\hat{V}\mathbb{I}_2\big]\,|s\rangle\,, \tag{22}$$

where $\vec{\sigma}$ is a vector whose components are the three Pauli matrices, q is the electric charge of the system, $\hat{A}$ is the electromangnetic vector potential operator, $\hat{V}$ the scalar potential operator, and $\mathbb{I}_2$ is the 2×2 identity matrix. Let the "lower-level" states be Dirac fermions

$$|\phi\rangle = N\int d^3k\Big[e^{iE(\vec{k})t}\big(c_1(\vec{k})\begin{pmatrix} (1+(1+\frac{k^2}{m^2})^{\frac{1}{2}})^{\frac{1}{2}} \\ 0 \\ \frac{k_3/m}{(1+(1+k^2/m^2)^{\frac{1}{2}})^{\frac{1}{2}}} \\ \frac{k_1/m+ik_2/m}{(1+(1+k^2/m^2)^{\frac{1}{2}})^{\frac{1}{2}}} \end{pmatrix} + c_2(\vec{k})\begin{pmatrix} 0 \\ (1+(1+\frac{k^2}{m^2})^{\frac{1}{2}})^{\frac{1}{2}} \\ \frac{k_1/m-ik_2/m}{(1+(1+k^2/m^2)^{\frac{1}{2}})^{\frac{1}{2}}} \\ \frac{-k_3/m}{(1+(1+k^2/m^2)^{\frac{1}{2}})^{\frac{1}{2}}} \end{pmatrix}\big) \tag{23}$$

$$+\,e^{-iE(\vec{k})t}\big(c_3(\vec{k})\begin{pmatrix} \frac{-k_1/m+ik_2/m}{(1+(1+k^2/m^2)^{\frac{1}{2}})^{\frac{1}{2}}} \\ \frac{k_3/m}{(1+(1+k^2/m^2)^{\frac{1}{2}})^{\frac{1}{2}}} \\ 0 \\ (1+(1+\frac{k^2}{m^2})^{\frac{1}{2}})^{\frac{1}{2}} \end{pmatrix} + c_4(\vec{k})\begin{pmatrix} \frac{k_3/m}{(1+(1+k^2/m^2)^{\frac{1}{2}})^{\frac{1}{2}}} \\ \frac{k_1/m+ik_2/m}{(1+(1+k^2/m^2)^{\frac{1}{2}})^{\frac{1}{2}}} \\ -(1+(1+\frac{k^2}{m^2})^{\frac{1}{2}})^{\frac{1}{2}} \\ 0 \end{pmatrix}\big)\Big]\,|k\rangle$$

where N is a normalization constant and c_{1-4} depend on the initial conditions. These states transition in accordance with:

$$\frac{\partial}{\partial t}\left|\phi\right\rangle = -i\big[\gamma^0(i\vec{\gamma}.\vec{\partial} - q\gamma^\mu A_\mu - m\mathbb{I}_4)\big]\left|\phi\right\rangle. \tag{24}$$

where γ^0 and $\vec{\gamma} = (\gamma^1, \gamma^2, \gamma^3)$ are the four Pauli matrices, $\vec{\partial} = (\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z})$, and $\mathbb{I}_4$ is the 4×4 identity matrix. Einstein notation has been used, which means repeating indices are summed over, with $\mu = 0, 1, 2, 3$. Expression (24) is essentially the Dirac equation, but rewritten to make it explicitly a set of temporal transition rules. Taking the Dirac theory as surrogate truth, one can robustly explain the success of spin states as follows:

1.1. Constraints: a) Only states of positive energy have interactive modes, which means one can effectively set $\forall k\ c_3(k) = c_4(k) = 0$. b) The spectrum of $\left|\phi\right\rangle$ quickly drops beyond a certain momentum $p_{\max}$, given by $c_1(k \geq p_{\max}) \ll 1$ and $c_2(k \geq p_{\max}) \ll 1$.

1.2. Resolution limits: Fluctuations of order $\frac{p_{\max}}{m}$ are ignored; that is: $O(\frac{p_{\max}}{m}) \approx 0$.

2. Mapping: Let $m(\left|\phi\right\rangle) = e^{-imt}P^+\left|\phi\right\rangle$, where P^+ is a projection operator from the space of four-spinors to the space of two-spinors in accordance with:

$$P^+\begin{pmatrix}\left|\psi_1\right\rangle\\ \left|\psi_2\right\rangle\\ \left|\psi_3\right\rangle\\ \left|\psi_4\right\rangle\end{pmatrix} = \begin{pmatrix}\left|\psi_1\right\rangle\\ \left|\psi_2\right\rangle\end{pmatrix} \tag{25}$$

3. Robustness: Using (24), one can prove that within the constraints and resolution limits above,

$$\frac{\partial}{\partial t}(e^{-imt}P^+\left|\phi\right\rangle) \approx -\frac{i}{\hbar}\big[\frac{1}{2m}(\vec{\sigma}.(\hat{p} - q\hat{A}))^2 + q\hat{V}\mathbb{I}_2\big](e^{-imt}P^+\left|\phi\right\rangle). \tag{26}$$

A proof of relation (26) can be found in (Holland and Brown [2003]). The robustness is evident from comparing (22) with (26).

4.6 Conclusion

In this chapter, I argued against the prevalent assumption that the only robust explanations of scientific success are referentially transparent explanations (truthlikeness, referential success, structural similarity), which I called the Fundamental Dichotomy. I argued that it is possible, at least in the case of dynamical theories, to provide robust explanations of the success of a false theory without appealing to referentially transparent notions. This type of explanation for dynamical theories is called tracking. Tracking provides a middle path between realism and antirealism, such that, in agreement with the realist, the theoretical terms are systematically related to unobservable reality so the theory could be said to be successful as a result of "latching on to something unobservable", but in agreement with the antirealist, this latching does not require representing unobservable reality in any transparent fashion.

5.0 Chapter 5: A Reinterpretation of *Umdeutung* in Prescriptive-Dynamical Terms

5.1 Introduction

In this chapter, I will argue that the history of the creation of modern quantum mechanics is a perfect lesson in the making of a prescriptive-dynamical theory. More specifically, the transition from Bohr to Heisenberg can be characterized as a struggle to break free from descriptive-ontological constraints and focus on the prescriptive-dynamical aspects in creating a theory.

In the following, I will begin Section 5.2 by introducing the puzzle of Heisenberg's cryptic reasoning in his groundbreaking paper of 1925 based on two vague principles of Correspondence and Observability. I will discuss some technical preliminaries in Section 5.3 that are needed for understanding the issues that motivated Heisenberg, then go on to discuss the Correspondence Principle (Section 5.4), the Observability Principle (Section 5.5), and finally how Heisenberg put the two together to make quantum theory into a fully prescriptive-dynamical theory (Section 5.6).

5.2 Heisenberg's Sketchy Principles

Heisenberg's renowned *Umdeutung* paper of 1925 ("A Quantum-theoretical Reinterpretation of Kinematic and Mechanical Relations") marks the advent of modern matrix mechanics. It is a truism among historians of quantum theory that Heisenberg was led to his "reinterpretation" of quantum theory by means of two principles: the Correspondence Principle (CP) as proposed by Bohr, and the Observability Principle (OP) as promoted by Pauli and Born. Heisenberg himself speaks of "sharpening the correspondence principle" (Heisenberg 1925a) and using "only relations between observable quantities" (1925b) as his guiding philosophies. However, both principles are notoriously vague and versatile, and that

has made it difficult to determine the exact role played by each principle in the development of Heisenberg's project. Some authors (Darrigol 1992, 273 ff.; Mehra and Rechenberg 1982, vol. II, 284 ff., Beller 1999) consider CP as the fundamental principle at work in Heisenberg's *Umdeutung* and chalk up Heisenberg's references to observability as reflections after the fact and mostly a matter of lip service to appease the two staunch advocates of Machian empiricism, namely Born and Pauli; other historians (Hendry 1984; Beller 1983; cf. Holton 1984, Ch. VII) find OP and the influence of Pauli to have played a much more central role. Apart from the obscurity of the two principles themselves, Heisenberg makes things more difficult by concealing much of his reasoning and mostly presenting the conclusions.[1]

My aim in this chapter is to argue that Heisenberg's utilization of both the Correspondence Principle and the Observability Principle serve the same purpose: to turn Bohr's quantum theory into a fully prescriptive-dynamical theory by removing the descriptive-ontological elements. As we shall see, in old quantum theory, CP served to provide the transition rules among Bohr's stationary states; these transition rules were to be computed through some sort of "averaging" of the Fourier coefficients that resolve the electron's initial and final orbits, and often also of the orbits in between. This meant that the transition rules were partially open-ended and partially fixed: they were open-ended insofar as the exact "averaging" procedure was left to be determined locally and empirically; but they were fixed, i.e. constrained, insofar as the transition amplitudes *had to* add up to a Fourier expansion of a classically allowed orbit.

However, much of the physics community at the time came to the conclusion that one must get rid of the orbits altogether. Heisenberg was the first to execute this in what shortly after came to be known as matrix mechanics. By removing the description of orbits from the theory, Heisenberg achieved two simultaneous goals: firstly he freed the transition amplitudes from descriptive (orbital) constraints, which eliminates the fixed part of the Correspondence Principle and leaves only the open-ended part, thus making the theory truly *prescriptive*; secondly he purged the theory of all quantities other than states and transition amplitudes, i.e. of all non-dynamical quantities (which he confusingly referred to as

[1]Heisenberg's omission of his reasoning includes both the philosophical and the mathematical arguments. But while the latter has been largely unpacked by subsequent readers (see Aitchison et. al. 2004), the former remains somewhat mysterious and is the subject of this chapter.

"unobservable quantities"), thus making the theory truly *dynamical*. Thus, what Heisenberg means by "sharpening the correspondence principle" essentially amounts to making CP fully open-ended, while what he means by using "only relations between observable quantities" amounts to making CP fully dynamical (both goals being achieved through removal of descriptive constraints). Therefore, the *prescriptive-dynamical* view is perfectly suited for understanding both of Heisenberg's principles as well as this episode of history more generally. As I've explained in the other chapters regarding the prescriptive-dynamical view, the descriptive aspects of theories are considered inessential or even detrimental to the predictive power of the theory, and the transition rules in particular should not be constrained by global descriptive requirements, as they are open-ended and local. And that is indeed the hard-earned lesson of Bohr and Heisenberg in the early 1920s.

In summary, I shall argue that Heisenberg's "reinterpretation" can be reinterpreted as the first direct attempt at constructing an explicitly and purely prescriptive-dynamical theory at the expense of the (inessential) descriptive-ontological aspects. Let us begin by studying some technical aspects of Bohr's framework that will prove crucial in our historical journey.

5.3 Preliminaries of Bohr's Atomic Theory

The story of Bohr's atomic theory begins in 1913 with Bohr's trilogy "On the Constitution of Atoms and Molecules", in which Bohr introduced the notion of discrete energy states known as *stationary states*, and ends in 1925 with Heisenberg's "A Quantum-theoretical Reinterpretation of Kinematic and Mechanical Relations", which is commonly taken as the birth of modern quantum mechanics. One crucial struggle that ran through this twelve-year journey, and bookended it by looming large in both of Bohr and Heisenberg's works, was the issue of determining and compactifying the probabilities (or amplitudes[2]) of transitions among various pairs of stationary states. Making the transition rules manageable is why CP

[2]A transition amplitude is a complex number whose modulus squared gives the transition probability: $|C_{nm}|^2 = P_{nm}$. One of Heisenberg's discoveries was that one needs the complex phase for the calculations to work out, and thus that solutions to the dynamical equations must always be transition amplitudes, not transition probabilities.

was introduced, and why Heisenberg invented his matrix mechanics (as a "sharpening" of CP). To see why transition rules were such a major source of grievance for quantum theorists between 1913 and 1925, we need to review some basics of Bohr's theory.

By the time Bohr entered the scene, experimentalists working on hydrogen and other elements had already found rigorous empirical regularities in the frequencies of spectral lines[3]. In particular, it was known that one could find a series of terms such that each observed frequency can be written as the *difference of two terms* in the series. These were called *term series* (Kragh 2012, 57-58). It was known, for instance, that the frequencies of hydrogen lines can be captured by the formula:

$$\omega_{nm} = R(\frac{1}{m^2} - \frac{1}{n^2}) \tag{27}$$

where R is the Rydberg constant and n and m are natural numbers and $n \geq m$. The series were typically given names for each choice of m: the case of $m = 1$ was called the Lyman series, $m = 2$ was referred to as the Balmer series, and so on.

Bohr's primary aim in his 1913 paper was to derive the Rydberg formula above. While his initial calculations follow a rather convoluted path, Bohr eventually provided an elegant way to deduce Rydberg's formula as follows. In any standard Kepler problem with the central force $F = -\frac{k}{r^2}$, one can show that

$$E = -\frac{k^2 m}{2J^2} \tag{28}$$

where E is the energy of an orbiting object of mass m, and J is the object's angular momentum. Given this, Bohr's innovation was essentially to make the following two assumptions:

$$J = nh \tag{29}$$

$$\omega_{nm} = \frac{1}{\hbar}(E_n - E_m) \tag{30}$$

where $\hbar = \frac{h}{2\pi}$ is Planck's constant and E_n and E_m are the energy contents of the initial and final orbits, respectively. The first assumption amounts to the claim that only a discrete subset of all classically allowed Kepler orbits are permitted in quantum theory, i.e. that the

[3]Line spectra are bands of light of discrete frequencies emitted when a dilute sample of an element is stimulated (usually through electricity).

electron can exist permanently only in a set of *stationary states*, while the second assumption amounts to stipulating that the electron emits no radiation unless it "jumps" from one stationary state to another, in which case it emits a monochromatic radiation carrying an amount of energy equal to the difference of energy between its initial and final states, which implies the frequency relation (30) once we recall Planck's relation $E = \hbar\omega$. The two formulae (29) and (30) together entail the Rydberg formula (27).

In this way, Bohr provided clear rules for assigning states to the atom and interpreted each spectral line as the result of a *transition* between two states. The fundamental assumption that allowed Bohr to provide state assignment rules was that the angular momentum of the electron must be an integer multiple of Planck's constant, i.e. $J = nh$. Since Bohr had originally assumed that electrons follow a two-dimensional orbit (staying on the same plane throughout), there was only one component of angular momentum to quantize. But what if there are more than one degrees of freedom, e.g. if the electron follows a three-dimensional trajectory? In that case, one would hope that a similar quantization condition can be imposed on every component of angular momentum. In Hamiltonian mechanics for periodic systems we have:

$$J_i = \oint p_i dq_i \tag{31}$$

where J_i is the ith component of angular momentum, q_i the ith generalized coordinate and p_i the corresponding conjugate momentum, and the integral is taken over a complete period. One would therefore hope for a quantization rule such as:

$$J_i = \oint p_i dq_i = n_i h \tag{32}$$

where n_i is the ith quantum number. However, this procedure is ambiguous, for with more than one degree of freedom, one can describe the system in any number of different coordinate systems (Cartesian, spherical, hyperbolic, etc.), and equation (32) would result in different quantizations depending on the choice of coordinates. So in what coordinate system should one quantize the angular momentum? The solution to this problem was provided by the action-angle formulation of Hamiltonian mechanics, which is designed to find the preferred coordinate system in which the total action is separable into a sum of actions for each

coordinate: $S(q_1, q_2, ..., q_d) = S_1(q_1) + S_2(q_2) + ... + S_d(q_d)$, where d is the number of degrees of freedom. The action-angle formalism was therefore perfect for Bohr's quantum theory.

However, the quantization condition (32) is still very limited in scope, since it only applies to exactly periodic orbits, and thus rules out many promising orbits that are not exactly periodic. The action-angle formalism came to the rescue again: another great advantage of this formalism is the fact that it allows one to decompose *non-periodic* orbits into a sum of terms, each of which is periodic, with each term corresponding to a distinct frequency. This was known as the theory of *multi-periodic* orbits, which remained at the foundation of old quantum theory until 1925. With the help of action-angle variables and the theory of multi-periodic orbits, one could write almost any orbit that remains confined to a finite region as a Fourier series:

$$x^{(n)}(t) = \sum_{\tau_1,\tau_2,...,\tau_d=0}^{\infty} R^{(n)}_{\tau_1,\tau_2,...,\tau_d} \cos[(\tau_1\omega_1 + \tau_2\omega_2 + ... + \tau_d\omega_d)t + \alpha^{(n)}_{\tau_1,\tau_2,...,\tau_d}] \tag{33}$$

$$= \sum_{\vec{\tau}=\vec{0}}^{\infty} R^{(n)}_{\vec{\tau}} \cos[(\vec{\tau}.\vec{\omega})t + \alpha^{(n)}_{\vec{\tau}}] \tag{34}$$

where the superscript (n) designates the nth stationary orbit, ω_i is the frequency corresponding to the ith preferred coordinate, the Rs are real numbers which may be function of the ωs themselves, and the αs are called *phases*. For the sake of simplicity, we will drop the vector notation from now on (or equivalently, we will focus on systems with one degree of freedom). We have:

$$x^{(n)}(t) = \sum_{\tau=0}^{\infty} R^{(n)}_{\tau} \cos[(\tau.\omega)t + \alpha^{(n)}_{\tau}] = \sum_{\tau=-\infty}^{\infty} C^{(n)}_{\tau} e^{-i(\tau.\omega)t} \tag{35}$$

where the C_τ are complex numbers given by:

$$C^{(n)}_{\tau} = R^{(n)}_{\tau} e^{i\alpha^{(n)}_{\tau}} \tag{36}$$

In order for $x(t)$ (or whatever function is being decomposed) to be real, we must have $C_{-\tau} = C^*_\tau$, where $*$ indicates complex conjugate. As we shall see shortly, decomposing the orbit into a Fourier series expressed in separable coordinates was crucial for the consistency and applicability of Bohr's theory and for the formulation of the Correspondence Principle.

Two aspects of Bohr's proposal were puzzling, if not shocking, to his contemporaries: first of all, Bohr's theory flew in the face of the longstanding empirical fact that the frequency of a wave, and in particular that of an electromagnetic wave, is identical to the mechanical frequency of oscillation at the source which generates the wave. A charged harmonic oscillator with a mechanical frequency ν, for instance, would generate electromagnetic radiation of frequency ν. Similarly, according to classical electromagnetism, a particle following an orbit such as equation (35) would simultaneously radiate waves of frequency $\tau.\omega$ for *all* values of τ for which the C_τ coefficients do not vanish. According to Bohr's theory, however, the electron emits *no radiation* while following $x^{(n)}$ above, and instead radiates only when jumping from $x^{(n)}$ to some lower orbit $x^{(m)}$ (assuming n is greater than m); the frequency of radiation emitted during this process has in principle *nothing* to do with the ω_i, seeing as it is given by the *difference of energy* between $x^{(n)}$ and $x^{(m)}$ in accordance with equation (30) above. The fact that Bohr's theory completely divorced the radiated frequency from the mechanical (orbital) frequency was not only surprising, but also potentially inconsistent with the results of classical electromagnetism.

The second puzzling fact about Bohr's theory was that it said nothing about the *intensities* of spectral lines: in any given spectrum, some lines appear brighter than others, suggesting that transitions corresponding to the brighter lines occur with higher probability than those of the fainter lines. Classically, the intensity of each mode of radiation of frequency $\tau.\omega$ is proportional to $|C_\tau|^2$ (Larmor's formula). But Bohr's original proposal provided no guidelines for how to approach the question of the regularities of intensities.

Now, from a prescriptive-dynamical point of view, the fact that Bohr's theory does not *predict* intensities is not only unproblematic, but actually desirable, because the prescriptive-dynamical view requires that transition rules ("LEMPs") be *open-ended* and determined *locally and empirically*. This means that the theory cannot possibly predict what the transition rules are going to be, and any theory that puts restrictions on the transition rules can get in trouble in new contexts of application. In Hamiltonian mechanics, for instance, the Hamiltonian encodes the transition rules, and the Hamiltonian is not a theoretically-fixed quantity: it simply depends on the system of interest. In diffusion theory, the diffusion function is not theoretically predicted or restricted, and is simply read off the local data

pertaining to the fluid under study (see Chapter 1 for elaboration of this point).

Nevertheless, a good dynamical theory provides a procedure for *compactifying* the transition rules. Hamiltonian mechanics encodes the transition rules in the Hamiltonian function, FLRW cosmology encodes them in the equation of state, and so on. Without a compactification technique, the transition rules will be a massive table of unorganized numbers. One could in principle read off the transition probabilities of different quantum jumps in Bohr's theory from the empirical data pertaining to the intensities of the lines, and these lines did in fact follow certain rather simple regularities such as $\Delta k = 0, \pm 1$ for the azimuthal quantum number in many cases. But there were too many exceptions and irregularities, and the procedure would often prove cumbersome and unsatisfactory. It was nonetheless practiced to some extent: for instance, Sommerfeld was in close collaboration with Paschen, whose experiments informed Sommerfeld's continual updating of his intensity rules (Kragh 1985; 2012, 162). But this was always a source of dissatisfaction with Bohr's framework.

Bohr attempted to get a handle on both troubling aspects of his theory by visiting a particular classical limit which I have referred to as the *low-frequency, high-quantum number* regime in Kaveh 2014. If the initial and final orbits are characterized by quantum numbers n and m, respectively, this regime amounts to the condition that $n - m \ll n, m$. First of all, it can be proven (Van Vleck 1924, ft. 4) that in this limit, the radiated frequency ω_{nm} is related to the mechanical frequency of motion ω through $\omega_{nm} \approx (n - m)\omega$. If we now let $\tau := n - m$, we recover the classical fact that the radiated frequencies are harmonics of the mechanical frequency as expressed in equation (35). This helps alleviate the first concern, for it shows that at least in some classical limits, the longstanding principle of the coincidence of mechanical and rediative frequencies can be salvaged. This is sometimes called the *correspondence theorem*, not to be confused with the Correspondence Principle.

At the same time, the correspondence theorem hints at a solution to the second puzzling aspect of Bohr's theory, namely the problem of intensities. If each term $\tau.\omega$ in the Fourier decomposition corresponds to a transition $n \to n - \tau$ with frequency $\omega(n, n - \tau)$, then one would expect each coefficient $C_\tau^{(n)}$ to be related to the *probability* $P_{n,n-\tau}$ of the transition $n \to n-\tau$. The reasoning is as follows. First of all, as said above, the coefficient $C_\tau^{(n)}$ provides the intensity amplitude for the mode of frequency $\tau.\omega$ in classical electromagnetism. Now,

classically, *all* modes are radiated simultaneously, whereas in Bohr's theory only one mode is emitted in each jump. However, if the probabilities of the jumps are adjusted appropriately, one would get the *appearance* of classical radiation from an *ensemble* of individual jumps. This implies that the transition probabilities of Bohr's theory must be related to the square of coefficients of the Fourier decomposition of the electron's orbit. The determination of the Fourier coefficients and their precise connection with transition probabilities drove much of the research in the following decade, and remained unsolved until the very end. As Born expressed one year before Heisenberg's *Umdeutung* in his "Über Quantenmechanik" (1924),

> The problem of the determination of the $[|C_\tau|^2]$ is closely related to the investigation into the intensities of spectral lines, and is of greatest importance for the further development of quantum theory. (Born 1924, quoted from van der Waerden 1967, 191)

The first problem was establishing a *unique, quantitative* relationship between the transition probabilities and the Fourier coefficients. In the low-frequency, high-quantum number regime, the initial and final orbits are very nearly the same, and their Fourier expansions are therefore nearly identical. This allows one to associate each mode of radiation with exactly one Fourier coefficient. In Bohr's theory, however, each mode of emission is associated with *two* orbits. The Fourier coefficients of which orbit should feature in the transition probability? A simple proposal would be to include the initial orbit only and simply let $P_{n,n-\tau} \propto |C_\tau^{(n)}|^2$, where $P_{n,n-\tau}$ denotes the probability of the jump $n \to n-\tau$, Although this prescription was used at times, there were various physical and metaphysical considerations that hinted at the possibility that the Fourier coefficients of the final orbit, and perhaps even those of the disallowed intermediary orbits, should also be taken into account.[4] In its most general form, then, the transition probabilities would be given by some weighted average of $C_\tau^{(i)}$, where $m \leq i \leq n$.

The averaging procedure was open-ended: for instance, Kramers, who devoted his dissertation to the question of *Intensities of Spectral Lines* (1919), suggested a logarithmic average, while Van Vleck (1924) recommended a polynomial average. There were two restrictions on the resulting transition probabilities, however: i) the formula for the average must be such that it goes over to a simple proportionality in the low-frequency, high-quantum number

[4] I shall not elaborate on these considerations. See Kaveh 2014, §2.

regime, thus recovering Larmor's radiation formula; ii) while the average of $C_\tau^{(i)}$ is taken "vertically" over $m \leq i \leq n$ for a fixed τ, the coefficients must also satisfy the "horizontal" requirement that each complete set of $C_\tau^{(i)}$ for $0 \leq \tau \leq \infty$ and fixed i specifies the ith *classically allowed orbit* (see Figure 3). As we shall see in §5.4, it was precisely the latter constraint that became the target of Heisenberg's attack.

The above considerations suggest the following prescriptions for "translating" classical quantities into their "corresponding" quantum mechanical ones:

$$\tau.\omega^{(n)} \rightarrow \omega(n, n-\tau) \tag{37}$$

$$C_\tau^{(n)} \rightarrow C(n, n-\tau) \tag{38}$$

where $C(n, n-\tau)$ is the quantum theoretical amplitude for the transition $n \rightarrow n-\tau$, or at least one of such amplitudes featuring in the grand average. (The arrows indicate substitution, not a mathematical or physical relation.) Moreover, following a series of promising results between 1924-1925, starting with Kramers's dispersion theory (1924a, 1924b) and Kramers and Heisenberg's follow-up paper (1925), and continuing with Born's (1924) and Born and Jordan's (1925) systematization of the Kramers-Heisenberg strategy, the following substitution recipe was proposed:

$$\tau \frac{\partial}{\partial J} F(J) \rightarrow \frac{1}{h}(F(n) - F(n-\tau)) \tag{39}$$

where J is the action variable. Prescription (39) is sometimes referred to as (Born's) *differential-to-difference rule*.[5] I will refer to prescriptions (37), (38), and (39) collectively as the *Born-Kramers-Heisenberg translation scheme*.

In the following section, I will summarize the correspondence arguments above in terms of a Correspondence Principle with four components, and argue that it was the orbital constraints (manifesting in Figure 3 above) that Heisenberg found problematic.

[5] A more intuitive but less useful formulation of the differential-to-difference rule could be obtained by letting $J = nh$ and $\tau = n - (n-\tau) = \Delta n$, which yields: $\frac{\partial F}{\partial n} \rightarrow \frac{\Delta F}{\Delta n}$. This is of course mathematically problematic, for n is not a continuous variable, and otherwise unhelpful because n does not feature in the classical theory.

5.4 Making the Correspondence Principle Prescriptive

The idea of the Correspondence Principle (CP) did not appear until Bohr's musings in 1918 and was not given its legendary title until Bohr's address to the German Physical Society in 1920. However, the germ of the idea was in the air from the very beginning and was used by Bohr and others in many different contexts. Future uses of the concept did not help with the ambiguity in meaning, as many physicists used the principle in ways that had very thin connections to Bohr's meaning (Kragh 2012, 219). Together with the inherent vagueness of the principle, this fragmentation in usage led to a widespread lack of clarity on what exactly CP was supposed to be. More specifically, one could identify four components to the Correspondence Principle:

> CP-i: *Statistics:* Intensities correspond to transition probabilities of *individual modes* in such a manner that in the appropriate limit, the statistical sum of these probabilities results in the same proportions as in classical radiation, in which all modes are emitted simultaneously.
> CP-ii: *Analogy:* The rules governing quantum transition probabilities follow a systematic analogy with the classical case.
> CP-iii: *Open-endedness:* While the state assignments are fixed, the exact form of the intensity formula is partially open-ended and must be determined empirically.
> CP-iv: *Orbital Constraint:* The intensities are partially fixed for they must be obtained from (some sort of averaging of) the Fourier coefficients that form the harmonics of a classical orbit.

The first component (*Statistics*), which Darrigol (1992, 220) calls "the narrowest form" of CP, was the most uncontroversial and secure one. The idea that intensities correspond to transition probabilities is a straightforward consequence of the assumption that only one unit of $\hbar\omega$ is emitted in each transition, and therefore that the intensity must indicate the *number* of times a unit of $\hbar\omega$ is emitted: the brighter lines must correspond to transitions that occur more often, i.e. with higher probability. Further, the assumption that the statistical sum of these probabilities in an ensemble must add up to the classical intensities is the only way to make Bohr's theory (in which each mode is emitted separately) compatible with classical electromagnetism (in which all modes are emitted simultaneously but in different proportions). As Bohr remarked:

> In the limiting region of large quantum numbers there is in no wise a question of a gradual diminution of the difference between the description by the quantum theory of the phe-

> nomena of radiation and the ideas of classical electrodynamics, but only of an asymptotic agreement of the statistical results. (Bohr 1924, 23)

Note that CP-i is *not* the simple requirement that the predictions of quantum theory be congruent with those of classical mechanics in the appropriate limit. Rather, it consists of two statements, one of which links intensities with transition probabilities and the other demands that the quantum-classical agreement in the limit be *statistical*, not exact. This is more specific than the generic condition that predictions of a successor theory go over to those of its predecessor in those domains in which the predecessor was successful. The latter condition, which I have called the *Congruence Requirement* elsewhere (see Kaveh 2014), was emphatically denied by Bohr as a component of CP: for instance, he is reported to have told Rosenfeld that "the requirement that the quantum theory should go over to the classical description for low modes of frequency is not at all a principle. It is an obvious requirement for the theory." (Rosenfeld, 1973, 252, quoted from Kragh, 2012, 197)

Considering the statistics in the low-frequency limit was often the starting point of correspondence arguments, and many physicists in this period (including Bohr and Heisenberg) occasionally used the phrase "Correspondence Principle" simply to refer to CP-i above. Sommerfeld, for instance, tended to use the term in this manner: in a 1924 letter to Kramers, he insisted that "the correspondence principle is a (highly valuable) limiting *theorem* of quantum theory, but not its *foundation*" (Mehra and Rechenberg 1982, vol. II, 155). Pauli was more explicit: "The correspondence principle demands only that these intensities [of classical and quantum theory] must become asymptotically equal for large k [i.e. large momentum]." (Mehra and Rechenberg 1982, vol. II, 158). Even Kramers and Heisenberg, who were most sympathetic to CP and understood it as more substantial than a mere limiting theorem still occasionally spoke of "satisfying the 'correspondence' requirement that the scattered radiation ... should coincide in the limit of large quantum numbers with ... classical theory" (Kramers and Heisenberg 1925, 707).

The second component, namely *Analogy*, was the most obscure one, for it was never entirely clear what was meant by an "analogy" between classical and quantum theory. Bohr often spoke of such analogies or "natural generalizations" of classical mechanics with a characteristic Bohrian ambiguity. For example, in a letter to Rutherford of 1917, a year

before his attempt at the first systematic development of the principles of quantum theory, Bohr wrote that an "analogy between this [quantum] theory and the ordinary theory of electrodynamics" has enabled him to "overlook the theory with all its different applications from a uniform point of view" (Kragh 2012, 191). Even as late as 1925, Bohr spoke of CP in this way:

> We must resort to symbolic analogies to an even higher degree than before. Just lately I have been racking my brain to dream up such analogies. (Bohr to Born 5/1/1925)

Physicists and physical chemists such as Reiche and Heurlinger who worked on applying Bohr's theory to molecular spectroscopy often made the "analogy principle" an axiom in their work (see Kragh 2012, 242). The analogy that Bohr himself seems to have chosen concerned the role of the Fourier coefficients in the two theories.[6] As we shall see below, Heisenberg's innovation was to relocate the analogy to the level of the *dynamical equations* governing the Fourier coefficients, as opposed to the solutions of said equations, namely the coefficients themselves.

The third component, i.e. *Open-endedness*, was the most implicit and amorphous aspect of CP, as it was rarely stated as a principle of its own; but in practice it was the source of some of the most fruitful applications of CP. As Darrigol explains:

> [T]he precise expression and the scope of the correspondence principle depended on the assumptions made about the electronic motion. Whenever this motion was a priori determined, the "correspondence" aided in *deducing* properties of emitted radiation. In the opposite case, characteristics of the electronic motion could be *induced* from the observed atomic spectra. This ambiguity made the correspondence principle a very flexible tool that was able to draw the most from the permanent inflow of empirical data. (Darrigol 1992, 83)

What Darrigol calls the "inductive side" of CP corresponds to the open-ended component: rather than making an a priori assumption about the orbit of the electron (which would then be decomposed into its Fourier harmonics to give the transition amplitudes), physicists often *reverse-engineered* the orbits from the empirical transition rules obtained in local experiments. This was especially helpful in the case of atoms with several electrons, whose

[6] Bohr later distanced himself from the "analogy" talk on the ground that "such an expression might cause misunderstanding, since ... the Correspondence Principle must be regarded purely as a law of the quantum theory" (Bohr 1924, 22).

Kepler problem yielded a formidable range of possible solutions, many of which were not (multi-)periodic. The open-ended aspect of CP allowed one to prune the set of possible orbits by appealing to the observed intensities and reading back the Fourier decomposition from them (see Darrigol 1992, 151).

This "reverse-engineering" of the quantities that encode transition rules from local empirical data perfectly fits the prescriptive-dynamical framework introduced in Chapter 1, according to which the state assignments are fixed, but the transition rules must be read back from the data by presupposing the state assignments and "fitting" the observed transition rules to them. For example, I argued in Chapter 1 that force formulae such as $F = -\frac{k}{x^2}$ and $F = -bv^2$, which encode the transition rules of Newtonian mechanics, are inexhaustible and therefore cannot be tested against data, but must be derived from them by presupposing $F = ma$ and measuring F empirically. One then plugs the reverse-engineered transition rules back into the theory to make predictions about the future behavior of the system. For instance, one puts together $F = -kx$ and $F = ma$ to get $\ddot{x} = -\frac{k}{m}x$, which provides final conditions for initial conditions of the form $(x, \dot{x})$. This is precisely the procedure followed in applying the open-ended aspect of CP. Here is Darrigol again:

> Bohr also used the correspondence principle in an inductive way, to infer features of the atomic motion from observed characteristics of spectra. This procedure, despite appearances, increased the predictive power of the theory, because the properties of motion induced from some spectral regularity could be used, in combination with a priori constraints on the motion, to deduce other phenomena, both physical and chemical. (Darrigol 1992, 283)

If my claims about the locality and open-endedness of transition rules are correct, then a full-blown prescriptive-dynamical theory would provide a way to *compactify* transition rules, but no way to *infer* them from a priori principles. This brings me to the fourth and last component.

The fourth component, *Orbital Constraint*, was the most specific and concrete aspect of CP, and it corresponds to Darrigol's "deductive side" of the principle, for it allows one to *deduce* the transition rules from assumptions about the shape of the orbit. One would begin by solving the classical Kepler problem, proceed to apply the Bohr-Sommerfeld quantization conditions (equation 32) in the appropriate action-angle coordinates, decompose the resulting orbits into their Fourier harmonics, and finally average the relevant coefficients of the orbits

between (and including) the initial and final states. As said above, the averaging procedure was left unspecified and was subject to speculation. In special cases, such as when all the $C_\tau^{(i)}$ corresponding to a given transition vanished, the exact form of averaging was unimportant and one could infer that the transition is forbidden. In most other cases, however, the averaging was a matter of guesswork.

The procedure was applied with mixed results. As the title *Intensities of Spectral Lines* makes clear, Kramers devoted his dissertation (1919) under Bohr to calculating the transition probabilities for the hydrogen atom especially in the case of fine structure and the Stark effect. Kramers's results were quite encouraging: not only did Kramers recover known facts about the relative intensities of lines in the case of fine structure and the Stark effect, he even made novel predictions about certain unexplored transitions in Stark effect experiments, which were later confirmed (Kragh 2012, 206). Similarly, in a coauthored paper of 1922, Sommerfeld and Heisenberg derived several selection rules and intensity ratios for multiplets and the Zeeman effect, also in good agreement with experiment (Sommerfeld and Heisenberg 1922).

Other applications of the "deductive side" of CP were not as successful. Frank Hoyt (1923; 1924; 1925) made copious calculations for the intensities of X-ray and Balmer series using several different averaging methods, but none of his results were even close to the experimental values (Kragh 2012, 214). Richard Tolman (1925) made similar attempts at calculating the probability of an electron falling into the nucleus, and got divergent results depending on the averaging method. Moreover, the application of CP to the Zeeman effect implied that each line must split into three finer lines (Darrigol 1992, 124). This was mostly but not always true. The splitting of certain lines into more than three components was dubbed the *anomalous Zeeman effect*, which remained a major obstacle for the success of Bohr's theory throughout. More generally, orbital constraints could only be applied to the problem of intensities in the case of multi-periodic systems, which as Darrigol points out, "covered hardly more than the hydrogen atom" (Darrigol 1992, 283). Dismayed by such unreliable success, Sommerfeld remarked:

> A correspondence treatment of the intensity problem yields only approximate values, in a way which does not seem to be very appropriate in view of the simplicity of the empirical facts. (Sommerfeld 1924, 1048)

The mixed success of the fourth component of CP is hardly surprising from a prescriptive-dynamical standpoint, for this component is the most problematic one if one accepts that transition rules are open-ended and should not be fixed in an a priori fashion. The "deductive" component violates open-endedness by imposing the requirement that the transition amplitudes must be obtained from coefficients that decompose a classical orbit. This is essentially what Figure 3 above illustrates: while the transition rules are calculated through a "vertical" average of the coefficients, the latter are also required to form a Fourier series of a classical orbit when arranged "horizontally". Therefore, the orbital constraints (CP-iv) partially fix the transition rules in an a priori manner. According to the prescriptive-dynamical view defended in this dissertation, this usually causes trouble for the theory down the line as soon as it is applied in localities in which those constraints are violated. Bohr introduced the orbital constraints to get a handle on *compactifying* the transition rules, but his strategy simultaneously prevents the transition rules from being fully local and empirical. If I am correct, the orbital constraints must be eliminated from Bohr's theory in order to make it a fully-fledged prescriptive-dynamical theory. And indeed, this is precisely what happened subsequently. As Darrigol explains:

> In the general process of freeing atomic motion from classical preconceptions, the deductive side of the correspondence principle shrank, until nothing seemed to be left of it, at least in the eyes of Bohr's most open critic, Wolfgang Pauli. (Darrigol 1992, 83)

"Freeing" the transition rules from orbital constraints is precisely what the prescriptive-dynamical view recommends to a theory such as Bohr's.

Heisenberg was the one to bring the "liberation" process to fruition. In a letter to Pauli (who shared Sommerfeld's skepticism about CP), Heisenberg effectively distinguished between CP-ii and CP-iv, embracing the former and denouncing the latter in one breath:

> [I]f one understands, as you, the correspondence principle to signify the wrong assertion that one may derive the quantum-theoretical intensity by averaging over the classical one, then *you* are right, and in that case one *cannot* obtain Ornstein's [empirical intensity] rule through correspondence principle; if, however, one understands it as signifying an analogous logical connection to the classical theory, then I am right. (Heisenberg to Pauli 10/8/1924)

The "wrong assertion" Heisenberg refers to is precisely the content of CP-iv, while the "analogous logical connection" is what is promised / demanded in CP-ii. The letter to Pauli

was written in 1924, a few months before Heisenberg's *Umdeutung* paper, when Heisenberg's ideas had not yet settled down. Years later, in an interview with Thomas Kuhn, a mature Heisenberg reflected on his thinking at the time as follows:

> So the whole thing was a program which one had consciously or unconsciously in one's mind. That is, how can we actually replace everywhere the orbits of the electrons by the Fourier components and thereby get into better touch with what happens? Well, that was the main idea of quantum mechanics later on. One could see, more and more clearly, that the reality were the Fourier components and not the orbits. The Fourier components were more real than the orbits and still somehow their connection was similar to their connection in classical mechanics. So one tried to look for those connections between the Fourier components which were true in classical mechanics and to see whether or not similar connections are also true in quantum mechanics... . So this was the whole trick. (Kuhn's 4th interview with Heisenberg, 1963)

Heisenberg's project, then, was to find a way to maintain the analogy with classical transition rules but remove the orbital constraints. This "sharpening" of CP would thus make the transition rules fully open-ended, and the Correspondence Principle fully *prescriptive*. We shall see in §5.6 how exactly Heisenberg carried this out.

Note Heisenberg's talk of the Fourier coefficients as "more real than the orbits". What should be considered "real" and thus legitimate to use in building a physical theory was another running theme for Pauli, Born, Heisenberg, and Bohr, and it constituted the second stream of thought that eventually flowed into Heisenberg's "reinterpretation" of quantum theory in 1925. This stream of thought was discussed under the rubric of the *Observability Principle*.

5.5 Making the Observability Principle Dynamical

The Observability Principle (OP) – advocated in particular by Pauli and Born in the context of quantum theory – was an influential idea in early 20th century to the effect that only "observable" quantities can figure in the formalism of a physical theory, i.e. that theories should be purged of all "unobservable" variables. What exactly was meant by "observable" was not as clear, however.

OP is often traced back to Ernst Mach, who was Pauli's godfather[7] and a major source of influence on him. Both Pauli and Born were also well-versed in general relativity and Klein's subsequent field theory, and often mentioned OP in that context, which was also under heavy Machian influence. Pauli's adherence to OP has also been linked to his encounters with Oppenheimer who, being Bridgeman's student, maintained a strong American operationalist attitude (Holton 1984, 173 ff.). The term "observability" is therefore commonly understood along positivist / operationalist lines, i.e. as referring to the property of being detectable by the five senses or at least being measurable. Under this reading, OP demands that any variable that does not stand for a collection of sensory inputs or measurements must be eliminated from theory.

The positivist / operationalist reading has been heavily criticized by various commentators (Beller 1999; Camilleri 2009; Lacki 2002; Wolff 2014; cf. Wüthrich 2016). I will shortly join this group of commentators to argue that the positivist reading is not at all how Heisenberg understood and applied OP, although my alternative interpretation of the principle will differ from the abovementioned commentators. Nevertheless, let us begin by recognizing that the positivist reading is not entirely unjustified. This is not only due to the personal and historical connections between Pauli and Born and major positivist / operationalist figures, but also because of various assertions made by them which hint in that direction.

One finds the most positivist-sounding characterizations of OP in Born's remarks, who would often link the idea to Einstein's relativity theory. For instance, in their co-authored paper "Zur Quantentheorie aperiodischer Vorgänge", Born and Jordan explain the idea as follows:

> A principle of great range and fruitfulness states that in the true laws of nature only those quantities enter which are in principle observable and determinable. (Footnote: Thus relativity theory came about as a result of Einstein's recognition of the in-principle impossibility of determining the absolute simultaneity of two spatially separated events.) When we apply this principle to optical problems, we arrive at the presumption [Vermutung] that the phase differences between the motion of electrons of one atom and those of another belong to these in-principle indeterminable quantities. (Born and Jordan 1925, 493)

To understand this passage, recall that classically, the $C_\tau^{(n)}$ are complex numbers given by equation (36): $C_\tau^{(n)} = R_\tau^{(n)} e^{i\alpha_\tau^{(n)}}$. The modulus of the complex amplitude $|C_\tau|^2 = R_\tau^2$ gives the

[7] See Mehra and Rechenberg 1982, vol. I, 377.

power radiated (i.e. intensity), while the α_τs specify the relative phases of the different modes of radiation. Since classically speaking all modes are emitted simultaneously, the relative phases are important, as they affect the composition of the radiation: the same two modes may be radiated in-phase or out-of-phase with each other, and those would be two different beams. Now, in the quantum case, one must use the Born-Kramers-Heisenberg translation schemes (37) and (38) to obtain the $C(n, n-\tau)$ for each mode of radiation of frequency $\omega(n, n-\tau)$. The quantum theoretical $C(n, n-\tau)$ will therefore also be complex numbers characterized by a modulus and a phase, but according to Born and Jordan the phase has lost its meaning in this case, for the different modes of radiation are *not* emitted simultaneously but rather one at a time. Since the different quantum modes are not superposed in a single beam, the relative phase between them seems irrelevant.[8] The only thing that matters is the *probability* of each individual transition, which is determined by the modulus squared of the coefficients, not by their phase. Born had also similarly declared in an earlier paper that he only considers transition probabilities $P(n, n-\tau) = |C(n, n-\tau)|^2$, not the amplitudes $C(n, n-\tau)$ themselves, to have "quantum-theoretical meaning" (Born 1924, 388, quoted from van der Waerden 1967, 191).[9]. On the basis of their observability principle and the claim that the phases are unobservable in the quantum case, Born and Jordan attempted to devise a theory that dealt directly with the transition probabilities, rather than amplitudes, which of course could not and did not succeed (for as we know from quantum mechanics, the phases are indispensable).

By linking the idea to Einstein's objections to absolute simultaneity, Born and Jordan give OP a very Machian feel. This Machian spirit seems to have been a Göttingen phenomenon. As Heisenberg later recalled:

> The idea of having a new theory in terms of observables did indeed originate in Göttingen and was closely connected with the interest in relativity theory that existed there. (Heisenberg, Conversations, 174, quoted from Mehra and Rechenberg 1982, 274)

Jordan also recalled discussions of observability in Göttingen to have been under Mach's influence (Jordan, AHQP Interview, Second Session, 31).

[8] One might think that the relative phases would matter if two systems encountered each other, but Born and Jordan argue that the relative phases could only be determined through an interference pattern, which does not happen for incoherent radiation from two different sources. See ibid.

[9] See also Born 1955, 258-259.

In any case, whatever the Göttingen gang might have meant by "observability", it is clear that Heisenberg did not adhere to it. An observability principle truly in the spirit of Born would have led to a theory in which transition probabilities (as opposed to the amplitudes) appear directly in calculations[10], which Heisenberg clearly violated. As far as amplitudes are concerned, Heisenberg was very clear:

> At first sight the phase contained in [the transition amplitude] would seem to be devoid of physical significance in quantum theory... . However, we shall see presently that also in quantum theory the phase has a definite significance that is analogous to its significance in classical theory. (Heisenberg 1925c, 264)

In any case, outside of Göttingen, the "Observability Principle" seemed to have nothing to do with observability per se, whether in its formulations by various physicists or its application by Heisenberg. Even Pauli, Mach's godson and the prophet of observability in quantum theory circles, does not seem to have used the term "observable" consistently in a positivist / operationalist sense. Pauli originally raised the observability principle as an objection to Einstein and Klein's field theories in particular, and all continuum theories in general. The most commonly quoted passage in this regard is the following:

> There is a physical-conceptual objection that should not be forgotten. In Weyl's theory we constantly operate with the field strength in the interior of the electron. For a physicist, however, the field strength is only defined as a force on a test body; and since there are no smaller test bodies than the electron itself, the concept of the electric field strength in a mathematical point seems to be an empty meaningless fiction. One should stick to introducing only those quantities in physics that are observable in principle. (Pauli 1919, 749-750)

This passage is fascinating. First of all, note that Pauli is not objecting to the use of field variables per se, but only to the practice of assigning values to them *in the interior of an elementary particle*. But the problem with this cannot be that humans are incapable of *observing* the field inside the electron, because as a matter of fact, in the epistemic / physiological sense of the word, the field is unobservable whether in the interior or the exterior of the electron. Fields are simply not the kind of thing that can be available to the human senses. Perhaps by "observable" Pauli simply means measurable, which would make the passage above more plausible. However, as Pauli himself explains, the field values must

[10]Not to mention that according to Born, OP implied that classical space and time themselves would have to be replaced by a "discrete" space (Beller 1983, 488).

be *inferred* from measurements, and so must other presumably "observable" quantities such as stationary states. But if any quantity that can *in principle* be inferred from experimental measurements is fair game, then electron orbits would arguably be "observable" as well, for as we saw above, the "inductive side" of CP was used precisely to that end.[11] Rather, Pauli seems to be concerned about quantities that are assigned *without an interaction that justifies the assignment*. Pauli makes this more explicit in a letter to Eddington:

> Indeed one can only define [the path of an electron] through the *action* of the electron on other elementary particles... . [T]he field concept only has a meaning when we specify a *reaction*... . (Pauli to Eddington 9/20/1923, quoted from Hendry 1984, 21 – emphasis added)

The hesitation thus seems to be about quantities that do not correspond to any interaction, i.e. a process of *change* of one thing under the influence of another. As I would like to say, Pauli is objecting to *non-dynamical* quantities which cannot be detected through the observation of a *transition*. This becomes even clearer once Pauli applies his Observability Principle specifically to the case of quantum theory: in a letter to Bohr, Pauli writes:

> I think that the energy and momentum values of the stationary states are something much more real than the "orbits". The (not yet reached) aim must be to deduce these and all other physical and real observable properties of the stationary states from the (integral) quantum numbers... . (Pauli to Bohr 12/31/1924)

Two crucial points can be gleaned from this passage: first of all, Pauli says he *thinks* that stationary states are real. This already shows that he cannot be under the impression that the stationary states (or the energy, momentum, etc.) are *observable* or directly *measurable*, for one would not need to *think* about the existence of observable things: if stationary states were observable, there would be no doubt about their reality. The fact that Pauli hedges his claim by using "I think" and "much more real" (as opposed to simply "real") indicates Pauli's full awareness of the fact that the stationary states themselves must be *inferred* from observations. Indeed – and this is the second point – this is confirmed in the very next sentence, where Pauli speaks of *deducing* "physical and real observable properties". Since observables need not be "deduced", the privilege given to the stationary states seems to be about something other than observability. As the previous passage indicates, this privilege

[11] Cf. Wolff 2014 for another argument against the "measurability" reading of OP.

comes down to the fact that the states are *interactive* or *dynamical*, while the orbits are ontological.

In fact, it is not clear to me that even Born himself used "observability" in the positivist sense. For again, the quantities that Born considers "observable", such as the energy of the stationary states, transition probabilities, and relative phases of the same radiation, must all be *inferred* from the data, sometimes through laborious calculations (see, e.g., Cassidy 1979, 202). And whenever he explicitly listed those quantities which he considered "observable", he would always list dynamical quantities. Here is an example:

> Besides frequencies, the emitted light possesses the observable properties of intensity, phase, and state of polarization, which are only approximately accounted for by the theory (§17). These exhaust the observable properties of the motion of the atomic system. However, our computation assigns additional properties to it, namely orbital frequencies and distances, that is, the course of motion in time. It seems that these quantities are, as a matter of principle, not accessible to observation. Therewith we arrive at the following judgement, that for the time being our procedure is just a formal computational scheme which, for certain cases, allows us to replace the still unknown quantum laws by computations on a classical basis. Of these true [wahren] laws we would have to require, that they only contain relations between observable quantities, that is, energy, light frequencies, intensities, and phases. (Born 1925, 113-114)

Perhaps one could argue that light frequencies and intensities are observable, seeing as frequency manifests itself in the color of the light and intensity in its brightness. This would already be a stretch, not only because the color and brightness are only truly observable for the visible spectrum, but also because observing color and brightness is not the same as observing frequency and intensity, which are highly theoretical concepts (an adherent of the particle theory of light would see the colors as well as the next person but would not associate them with frequencies). But let us grant that frequencies and intensities of light are observable. We are still left with the puzzling statement that the energy and phase of the stationary states are also observable. This is clearly false, for the energies are given by Bohr's frequency condition (eq. 30): $\omega_{nm} = \frac{1}{\hbar}(E_n - E_m)$, in which ω_{nm} is the only measurable (or "observable") quantity; E_n and E_m must be inferred from those measurements. As for the observability of phases, Born is simply changing his verdict from a year ago, which perhaps suggests an evolution of the concept of observability during this period.

Other passages in which reference to OP is made confirm the theory that "observable

quantities" is best replaced by "dynamical quantities". For example, Kramers prided himself on the purity of his derivation of his dispersion formula (which proved to be a stepping stone to Heisenberg's *Umdeutung*) with the following words:

> [I]t contains only such quantities as allow of a direct physical interpretation on the basis of the fundamental postulates of the quantum theory of spectra and atomic constitution, and exhibits no further reminiscence of the mathematical theory of multiple periodic systems. (Kramers 1924b, 311)

What quantities appear in "the fundamental postulates of the quantum theory of spectra and atomic constitution"? Stationary states and transition probabilities. Thus, Kramers is in essence saying that his theory contains nothing but dynamical variables.[12] Note, by the way, that Kramers associated the derivation of his formula with the recent Bohr-Kramers-Slater theory (1924), which spoke of many truly unobservable quantities such as "virtual oscillators" and "virtual fields".

Finally, Heisenberg himself used OP in a similar manner, i.e. as the principle that only dynamical variables (states and transitions) can figure in a theory. For instance, he wrote to Pauli about his conviction that

> In calculating any quantities, like energy, frequency, etc., only relations between those quantities should occur, which can be controlled in principle. (Heisenberg to Pauli 6/24/1925)

Heisenberg's use of "control" is intriguing. Wolff (2014) argues partly on the basis of this passage that what Heisenberg really meant by "observability" was "causal efficacy", where "causal" is understood in a manipulationist / interventionist manner. This does not seem quite right, however. First of all, the orbits are causally efficacious in Bohr's theory in the sense that the Fourier coefficients of the orbit determine which transitions will and will not occur. Secondly, since in Bohr's theory the states correspond to orbits in a one-to-one manner, insofar as the states are casually efficacious, then so must be the orbits. Therefore, it would be odd if "causal efficacy" ruled out orbits but not states. It is true, however, that what one *directly controls*, or as I would like to put it, what one directly *interfaces* with (prepares, detects), are the states. In other words, the prescriptions of the theory are prescriptions for preparing and detecting the system in a particular *state*, not in a particular

[12] See also Kramers and Heisenberg 1925, in van der Waerden 1967, 234 for virtually identical statements.

orbit. To be sure, as said above states and orbits correspond to each other one to one, but the *human interface* is with the states, not with the orbits. One prepares / detects the system in a particular state, and then infers the orbit from what one assumes the prepared / detected state to be. Once again, I suggest that we make sense of things by letting "controllable" mean "dynamical".

Incidentally, the idea of a human interface also explains why the term "observability" would have been used. While both states and orbits must be inferred from more basic observations / measurements, the nodes at which experimental interface occurs are the states. In other words, the orbits are as it were *two steps removed* from measurement: one infers (prepares, detects) the states based on direct measurements, then records the transition rules among the states, and only then infers the orbits on the assumption that the transition amplitudes add up to a Fourier decomposition of an orbit. Put differently, while dynamical variables are not observable / measurable themselves, they are *the closest* point of interface with the system to observation / measurement.

The requirement that only dynamical quantities (states and transitions) be allowed is consistent with Heisenberg's occasional pronouncements that all physical quantities of interest can be deduced from transition amplitudes. Shortly before arriving at his "reinterpretation", Heisenberg wrote to Kronig:

> In classical theory the knowledge of the Fourier series for the motion is sufficient to calculate *everything*.
> [A]lso in the quantum theory everything is determined by the knowledge of the transition probabilities or the corresponding amplitudes.
> What I like in this scheme is that one can really reduce all interactions between atoms and the external world ... to transition probabilities. (Heisenberg to Kronig 6/5/1925)

Therefore, I conclude, one must interpret Heisenberg's Observability Principle as a "Dynamicity Principle", i.e. the requirement that a theory only involve states and transitions, and dispense with ontological constructs that must be inferred from the dynamical quantities.

In the final section, I will combine all of the considerations above to present a prescriptive-dynamical reinterpretation of Heisenberg's *Umdeutung*.

5.6 *Umdeutung:* the Making of a Prescriptive-dynamical Theory

In this section, I will present my prescriptive-dynamical reinterpretation of Heisenberg's groundbreaking paper "A Quantum-theoretical Reinterpretation (Umdeutung) of Kinematic and Mechanical Relations". My exegesis does not add anything to our current understanding of the formal / mathematical aspects of Heisenberg's work[13], but merely revisits them in light of the arguments advanced above.

To begin, I propose that the following principles are at work in Heisenberg's *Umdeutung*:

> (I) Only dynamical quantities (i.e. states and transition amplitudes) can feature in the theory (the "Observability" Principle).
> (II) The quantum theoretical frequencies and transition amplitudes correspond to their classical counterparts in accordance with the Born-Kramers-Heisenberg translation scheme (relations (37), (38), and (39)) in such a way that in the appropriate limit, the statistical sum of the amplitudes generates the classical spectrum.[14]
> (III) The transition amplitudes satisfy *analogous dynamical equations* to their classical counterparts.[15]
> (IV) The Bohr-Sommerfeld quantization condition (eq. (32)) appropriately translated in accordance with the Born-Kramers-Heisenberg scheme.
> (V) The transition amplitudes are subject to no other constraints than the ones above.[16]

Note that (I), (III), and (V) are all closely related to the idea of removing the orbits. By removing the orbits, one purges the theory of its descriptive-ontological aspects, allowing one to focus on the dynamical state-assignment and transition rules, which gives us (I). Moreover, prior to *Umdeutung*, the translation schemes were applied to the *solutions* of the classical equations of motion, namely the Fourier coefficients of classical orbits. By removing the orbits, Heisenberg arrived at the innovative idea that the translation scheme must be applied to the *equations of motion themselves*, rather than their solutions, and this gives us (III). As he reflected later on:

> One could see that the Fourier components were the reality, and not the orbits. So one had to look for those connections between the Fourier components, which were true in classical mechanics, and to see whether or not similar connections were trule also in quantum mechanics – i.e. if one took, instead of the Fourier components, the transition amplitudes of the

[13] See Aitchison et. al. 2004 for an excellent reconstruction.
[14] This is an augmented version of CP-i (*Statistics*).
[15] This is a special case of CP-ii (*Analogy*).
[16] This is a sharpening of CP-iii (*Open-endedness*).

real lines of atoms. (Heisenberg, Conversations, 135; quoted from Mehra and Rechenberg 1982, vol. II, 189-190)

Finally, by removing the orbits, Heisenberg eliminated the problematic component of CP, namely the orbital constraints (CP-iv) which prevented the transition rules from being maximally open-ended (or prescriptive), which gives us (V). Hence my claim above that Heisenberg achieved both aims of making the theory fully prescriptive and making it fully dynamical by removing the orbits. As he wrote to Pauli, "all my wretched efforts are devoted to killing totally the concept of an orbit" (Heisenberg to Pauli 7/9/1925).

Heisenberg begins by listing various classical relations and seeks after their quantum-theoretical translations. First of all, Bohr's frequency condition (eq. (30)) can be derived from the classical relation

$$\omega_{\tau}^{(n)} = \tau \frac{dE^{(n)}}{dJ} \tag{40}$$

where E is the energy written as a function of the action variables J. According to the Born-Kramers-Heienberg translation scheme (39), the quantum-theoretical version would be:

$$\omega(n, n-\tau) = \frac{1}{h}(E_n - E_{n-\tau}) \tag{41}$$

which is just Bohr's frequency condition. This in turn implies what is known as the Ritz combination rule:

$$\omega(n, n-\tau) + \omega(n-\tau, n-\tau-\sigma) = \omega(n, n-\tau-\sigma) \tag{42}$$

Next Heisenberg proceeds to apply (II) to replace each classical term $C_{\tau}^{(n)} e^{i\tau\omega^{(n)}t}$ with its quantum-theoretical counterpart $C(n, n-\tau)e^{i\omega(n,n-\tau)t}$. However, Heisenberg goes on to point out, one must resist the temptation to form a "Fourier series" in terms of a sum of quantities $C(n, n-\tau)e^{i\omega(n,n-\tau)t}$.[17] In light of (I) and (V), this is just as well.

[17] Heisenberg and Kramers (1925) attempted this strategy and failed.

Next, Heisenberg sets out to address the following problem. We know that given a classical orbit $x^{(n)}(t)$ characterized by Fourier coefficients $C_\tau^{(n)}$, one can form the Fourier series of $x^2(t)$ with coefficients $D_\tau^{(n)}$, where we have:

$$D_\tau^{(n)} e^{i\tau\omega^{(n)}t} = \sum_\sigma C_\sigma^{(n)} C_{\tau-\sigma}^{(n)} e^{i\tau\omega^{(n)}t} \tag{43}$$

What is the quantum-theoretical counterpart of this relation? Heisenberg suggests the following "simplest and most natural assumption"[18]:

$$D(n, n-\tau)e^{i\omega(n,n-\tau)t} = \sum_\sigma C(n, n-\sigma)C(n-\sigma, n-\tau)e^{i\omega(n,n-\tau)t}. \tag{45}$$

This is of course simply matrix multiplication, as Heisenberg, Born, and Jordan realized soon after. Heisenberg comments that despite first appearances (or, to read between the lines, despite Born's arguments to the contrary), the phase is indeed important, because the phases mix in expressions such as above, and therefore their relative values matter.

After the above kinematical considerations, Heisenberg then proceeds to examine dynamics. To deviate slightly from his order of presentation, let us consider his treatment of the Bohr-Sommerfeld quantization condition in accordance with (IV) above. First of all, from the Fourier expansion of $x(t)$ and some basic calculus we obtain:

$$J = \oint p dq = \oint p\dot{q} dt = \oint m\dot{x}^2 dt = 2\pi m \sum_\sigma |C_\sigma^{(n)}|^2 \sigma^2 \omega^{(n)}. \tag{46}$$

It is not possible to apply the Born-Kramers-Heisenberg scheme to this formula directly, so Heisenberg employs the following neat trick: differentiate both sides with respect to J to obtain:

$$1 = (2\pi m) \sum_\sigma \sigma \frac{d}{dJ}(\sigma\omega^{(n)}|C_\sigma^{(n)}|^2) \tag{47}$$

[18]In fact, this relation is not at all the simplest and most natural assumption that would follow from an application of the Born-Kramers-Heisenberg translation scheme to equation (43). Rather, one would expect to have:

$$D(n, n-\tau)e^{i\omega(n,n-\tau)t} = \sum_\sigma C(n, n-\sigma)C(n, n-(\tau-\sigma))e^{i\omega(n,n-\tau)t}. \tag{44}$$

Heisenberg is omitting an argument based on the requirement of the Ritz combination rule (42) that would justify replacing the relation above with (45). For a reconstruction of this argument see: Aitchison et. al. 2004

Now we have expressions of the form $\sigma\frac{d}{dJ}$, $\sigma\omega^{(n)}$, and $C_\sigma^{(n)}$, all of which we know how to translate in accordance with (37), (38), and (39). We get:

$$1 = \frac{4\pi m}{h}\sum_{\sigma>0}\left[|C(n,n+\sigma)|^2\omega(n,n+\sigma) - |C(n,n-\sigma)|^2\omega(n,n-\sigma)\right] \tag{48}$$

As Jordan proved shorly afterwards, (48) is just the canonical commutation relation $[q,p] = i\hbar$ where q and p are matrices (Darrigol 1922, 276 ff.). This is the relation used in the modern formulation of quantum mechanics.

Now, Heisenberg points out that up to this point, the Correspondence Principle had been applied by first *solving* an equation of motion of the form $\ddot{x} + f(x) = 0$ in action-angle variables and then applying the Bohr-Sommerfeld quantization conditions. This violates (I) and (V), for it imposes orbital constraints on the transition amplitudes: the solutions to a classical equation of motion are classical orbits, and requiring that the quantum transition amplitudes add up to an orbit violates the principle that the theory must involve only dynamical quantities (of which a classical orbit is not one) as well as the principle that the transition rules be open-ended beyond the requirements of (I)-(IV). Thus, in place of the traditional procedure, Heisenberg suggests following (III), i.e. applying the translation scheme directly to the equation of motion. At this point, however, Heisenberg has not yet equipped himself with matrix calculus, which makes it difficult to produce an explicit quantum-theoretical counterpart for the equation of motion, seeing as one is not allowed to form a "Fourier series" of the quantum-theoretical quantities to plug into the equation of motion. In light of this, Heisenberg suggests that we extract from the equation of motion *recursive relations* between the classical Fourier coefficients, and then translate these recursive relations into quantum theory using the Born-Kramers-Heisenberg scheme. The rest of Heisenberg's paper is devoted to illustrating this procedure using simple examples of aharmonic oscillators, which we will not get into. In modern quantum mechanics, the recursive relations are replaced by matrix equations that imitate classical mechanical equations of motion but with the dynamical variables replaced with matrices. Thus Heisenberg invented modern quantum mechanics in the form of matrix mechanics, and did so by focusing on the construction of a purely prescriptive-dynamical theory.

Note, by the way, that the relations obtained by Heisenberg are still not sufficient for fixing the transition amplitudes except in certain special cases (Darrigol 1992, 268), which testifies to their open-endedness in accordance with (V) above.

5.7 Conclusion

I have argued that Heisenberg's *Umdeutung* as well as the role of both principles of Correspondence and Observability in its formation can be fully understood by taking Heisenberg's "sharpening" of CP as the attempt to make the latter principle fully prescriptive (i.e. open-ended) and his promise to use only "observable variables" as the requirement that only dynamical variables (states and transitions) appear in the theory. As we have seen, both of those goals were intimately connected with the removal of orbits from Bohr's theory.

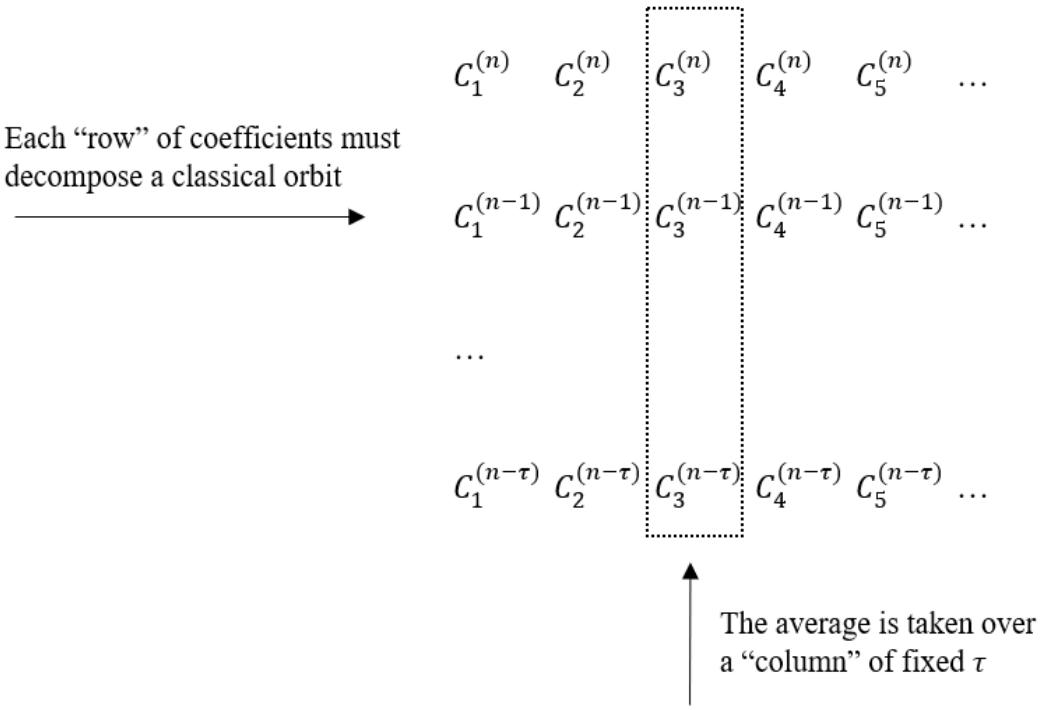

Figure 3: Transition probabilities constrained by orbital requirements

Appendix A On Productive and Unproductive State Assignments

In §2.4.2.1, we introduced the idea of productive and unproductive theories. The former connect to experiment at an early stage through a fixed set of instructions for assigning dynamical states to the system, and thus allow us to construct new LEMPs for every new system we come across to which our state assignments are applicable. Unproductive theories, on the other hand, have no fixed state assignment rules and must derive their precise form of state assignment *after the transition rules are known*. This means that upon experimenting with a new system, such a theory would have no recipe for constructing a new local set of transitions rules (LEMPs). Unproductive theories can only connect to experimental systems by piggy-backing on a productive theory which has allowed us to record the required transition rules. In this Appendix, I will elaborate on all of this in more technical detail. As said, this section is intimately connected to the material discussed in §2.4.2.1.

Experimental state assignment is a highly non-trivial and historically rich process. One might use a "ladder" structure: begin by assigning the quantity of interest to some unique measurement procedure that is considered authoritative for determining the value of said quantity, and then through a series of inductive steps establish what other measurements happen to track the same quantity, and eventually build an experimental repertoire for assigning values to the quantity. Measures of x are some of the oldest inventions in history and there is much to say about them, but the complexity of the situation is more readily apparent in the cases of $\dot{x}$ and p. In the former case, one might begin with a measurement procedure that is as faithful to the mathematical definition of $\dot{x}$ as possible, i.e. by measuring two positions very close to each other (e.g. using photogates or reconstructed trajectories) and dividing their separation by the time it took the object to travel from one to the other. Once $\dot{x}$ has been assigned to this procedure, one can then go on to show through a series of inductive steps that measurements based on the Doppler effect and other methods return the same values in the contexts of interest. Any of the procedures in this fixed experimental repertoire will then count as legitimate assignment (detection, preparation) of $\dot{x}$.[1]

[1] As always, "fixed" means *fixed across the applications of the theory*, not *across time*: the experimental

What about the Hamiltonian momentum p? Can it be measured *independently*, i.e. without going through a calculation mediated by velocity? The answer to whether or not Hamiltonian mechanics can be an autonomous, productive theory depends on the answer to this question. We need at least one measurement procedure that we can assign to p so that we can build a repertoire of procedures on its basis. The most natural candidate for such a procedure would be measurements of surface impact and cross-section in collision, because those quantities depend on momentum, not on velocity. If it is accurate to say that collision experiments measure momentum without going through a measurement of velocity, then this state assignment makes Hamiltonian mechanics an autonomous theory which does not piggy-back on Lagrangian mechanics. States of the form (x, p) can be assigned independently of transition rules and the theory is therefore productive.

The question then becomes: what is the relationship between the productive Hamiltonian mechanics and the parasitic one that is obtained from the Lagrangian solution through the Legendre transformation? It seems to be an empirical fact that in many ordinary cases, the two coincide. But this is not always the case. There are conjugate momenta derivable from the Legendre transformations that are *not* the momentum of the system. Let me illustrate this with an example.[2] Consider a linearly damped system (e.g. a particle moving with moderate velocity through air at zero gravity) with the EOM:

$$\ddot{q} + b\dot{q} = 0 \tag{49}$$

where b is some constant. This EOM can be compactified into the following (admittedly quirkly) Lagrangian:

$$L = \dot{q} \ln \dot{q} - bq \tag{50}$$

repertoire is surely being expanded through new measurement techniques all the time, but at any given time, it is the same repertoire regardless of local context.

[2]Some other aspects of this example are discussed by Smith (2008, 335) in the context of Noether's theorem.

When plugged into the Euler-Lagrange equations (the master formula), this LEMP will result in the EOM above. Now consider a Legendre transformation of the Lagrangian. We have:

$$p = \frac{\partial L}{\partial \dot{q}} = \ln \dot{q} + 1 \tag{51}$$

$$H = p\dot{q} - L = \dot{q} + bq \tag{52}$$

Plugging in the above Hamiltonian in Hamilton's equations will surely return the same equation of motion as before. This is indeed a model of Hamiltonian mechanics and it does predict the behavior of the linearly damped system.

However, the Hamiltonian $H = \dot{q} + bq$ is *not* the energy of the system. Here is why: H is derived from a Lagrangian that is time-translation invariant, which means that H is conserved as a result of Noether's theorem; but the actual energy of the system is not conserved, for it is damped. Therefore, H is not the energy of the system, which in turn implies that the corresponding canonical momentum p is not what one would obtain in impact and cross-section measurements. Therefore, the parasitic Hamiltonian theory cannot relate its state assignment rules to a fixed repertoire of experimental procedures, which is to say that it violates EPOC and is therefore an unproductive theory. In other words, while it is true that *given* the above formulae for momentum and Hamiltonian, one would get back the desired EOM from Hamilton's equations, the Hamiltonian in question could not have been discovered directly through local empirical measurement. It must necessarily be embezzled from the Lagrangian framework. Note that my point is *not* that this Hamiltonian model is unphysical: it is indeed as physical as a particle moving in air with moderate speed at zero gravity. The problem is rather that this Hamiltonian model can only exist as a reformulation of the Lagrangian model; it cannot stand on its own.

Note that the autonomous Hamiltonian mechanics which attaches its momentum assignments to fixed experimental procedures could not have handled this case. Since the energy is time-dependent, a linearly damped system simply has no model in the autonomous Hamiltonian framework. Consequently, the autonomous Hamiltonian theory has a narrower scope of application: there are systems that the Lagrangian theory and the parasitic Hamiltonian formalism can handle but the autonomous Hamiltonian cannot.

This brings us to a potential worry about the use of Hamiltonian mechanics. Given the above, the parasitic "Hamiltonian theory" might seem quite pointless: if the aim is to arrive at an equation of motion, why bother converting things into Hamiltonian form only to convert them back later? It would seem that one always ends up eliminating the canonical momentum from the two Hamilton equations to get back the same equation of motion that one could have derived from the Euler-Lagrange equations to begin with. Yet this procedure seems quite routine in a physics textbook. The answer is that physicists introduce this framework merely as a computational tool in Lagrangian mechanics for cases in which numerical and simulated solutions are called for, as well as cases in which there are cyclic coordinates, for which the Hamiltonian formulation is simpler. As Goldstein explains:

> It has been remarked that the Hamiltonian formulation is not particularly helpful in direct solution of mechanical problems. Often one can solve the first order equations only by eliminating some of the variables, e.g., the p variables, which speedily leads back to the second order Lagrangian equations of motion. But an important exception should be noted. The Hamiltonian procedure is especially adapted to the treatment of problems involving cyclic coordinates. (Goldstein 1980 [1950], 351)

Finn concurs in his *Classical Mechanics*:

> The Hamiltonian method is better suited for numerical calculations and for phase space visualizations. However, unless there are cyclic coordinates, the direct use of Hamilton's equations does not significantly reduce the complexity of the problem from an analytical point of view. (Finn 2008, 130)

The fact that the parasitic Hamiltonian is not a theory of its own but rather a handy side-note to Lagrangian mechanics is also obvious from the physicist's common practice of forming "Routhian" functions (Goldstein 1980 [1950], 352; Finn 2008, 137). Routhians are hybrids between Lagrangians and Hamiltonians, and their purpose is to allow us to treat only the cyclic coordinates in a Hamiltonian fashion, for there is no reason to bother with Legendre-transforming the non-cyclic coordinates.

Appendix B Evidence that Bohr's Failure Was a Failure of Tracking

The aim of this appendix is to justify a claim I made in §3.3.1.3, namely that the main failure of Bohr's model with regards to helium was its inability to accommodate the no combination rule (the fact that orthohelium and parahelium states never transition among themselves), which is a failure of tracking.

As said above, orthohelium and parahelium were originally considered separate elements due to the fact that their term series never mixed: one never observes spectral lines that indicate a transition from a para-state to an ortho-state or vice versa. However, a series of fascinating and meticulous experiments by Frank, Hertz, Reiche, and others soon established that the "ground state" of orthohelium is in fact an excited state, equivalent to one electron being in $1s$ and the other in $2s$ or $2p$, but for some reason *this* $2p$ does not decay spontaneously: it is a "meta-stable" state. So the two series of spectral lines turned out to be parallel series for the same element, except one of them has no Lyman series ($x \to 1s$). These experiments left no doubt that $2s_p$ is distinct from $2s_o$ (Gearhart 2017).

This immediately raised a puzzle: why the lack of mixing, given that all of these states are energy layers of the same atom? What prevents transitions from parahelium states to orthohelium states and vice versa? It was understood that any quantum theory of the helium atom would have to answer this curious question or be doomed.

Around the same time, Landé proposed a Bohr-style model in which helium's two electrons could either occupy coplanar or perpendicular orbits, and hypothesized that the former corresponds to orthohelium and the latter to parahelium.[1] Bohr was encouraged by Landé's work but wanted to find a secure foundation for it in the principles of his own theory. According to the latest version of Bohr's theory, electrons could occupy multiply periodic orbits[2], which gave the theory much more flexibility and crucially allowed for orbits that "penetrate"

[1]This was also taken to explain the fact that parahelium lines are singlets whereas orthohelium lines are doublets (as said we know now that they are in fact triplets!), for the coplanar electrons could revolve either in the same or opposite direction to each other. See Van Vleck 1926, 87.

[2]These are orbits that can be solved through separation of variables in the Hamilton-Jacobi theory, and thus can be decomposed into an (infinite) sum of periodic terms. While such orbits are in general not strictly periodic, given sufficient time they come arbitrarily close to the original position and velocity

under the inner layers. Using this model, Bohr and Kramers offered a variant of Landé's model in which the parahelium electrons orbited at a relative 120 degree angle (Darrigol 1992, 164).[3]

And this all seemed fine and good at first. Frank and his colleagues believed that they had tested and (somewhat qualitatively) confirmed Bohr's model through a series of experiments on helium ionization (e.g. Franck and Einsporn 1920). To be sure, there were quantitative discrepancies between the experimental values of helium ionization energy and those predicted from the Bohr-style model, but this was not taken as a definitive refutation of Bohr's theory, as the issue was widely considered a matter of finding the right orbital configuration.[4] Bohr himself tried to explain away the quantitative discrepancies by arguing that the ground state of helium is unstable, which according to Bohr results in "indeterminacy" ("Unbestimmtheit") in its energy (Darrigol 1992, 176). Kramers similarly rationalized that the theoretically obtained energies should not be identified with ionization energies to begin with, because the latter depend not just on the final state, but on the *process* through which the second electron arrives at its bound orbit from infinity (Kramers 1923, 340-341). As late as 1922, a year before it was considered refuted, Born declared the Bohr-Kramers model "the most probable configuration of helium" (Born 1922, 677).

So what was the real failure of Bohr's model? The answer lies in a set of state assignment rules that Bohr introduced in his 1918 "On the quantum theory of line-spectra" as a generalization and clarification of the old quantization rules $\oint pdq = nh$. One of the fundamental principles that Bohr (1918) drew on to assign stationary states was called the "Principle of Mechanical Transformability"[5], which Bohr considered necessary both for well-defined energy assignments and for his Correspondence Principle to be consistent. Without getting into much detail, the principle states that given a quantum mechanically allowed stationary

[3]Edwin Kemble proposed a similar model in 1921. The model is sometimes called the Bohr-Kramers-Kemble model. See Kragh 2012, 253.

[4]See, e.g., Sommerfeld 1923 for an example of attempts at solving the ionization energy problem. See also Van Vleck 1926, 86 ff. for a survey of the different orbital models of helium and how they fare against experimental values. Van Vleck presents the quantitative mismatch as a problem for helium models (ibid, 87 ff.), but places it second to Born and Heisenberg's 1923 proof – which I will shortly argue was the truly devastating problem – on the list of objections (ibid, 105-106).

[5]Van Vleck calls it the "Principle of Continuous Formation" (Van Vleck 1926, 98). Bohr introduced this principle having been inspired by Ehrenfest's adiabatic theorem; see Navarro and Perez 2006 for a discussion of the genesis of Ehrenfest's theorem.

state, all and only those orbits are allowed that can be connected to said state through a continuous, multi-periodic, classical path (see Perez 2009; Perez and Valls 2015). This principle in general gives rise to several alternative families[6] of orbits, each family closed under mechanical transformability, such that states in each family can transitions among themselves, but not across other families.

Now, the crucial point is this: Bohr and Kramers originally thought that their models of coplanar and crossed orbits belong to different families prescribed by the Principle of Mechanical Transformability. Since the two sets of states belong to alternative quantization schemes, they could not be realized in the same atom at the same time. At any given moment, the energy levels of the helium atom are quantized in accordance with either one or the other family of state assignments, but not both. Therefore, Bohr and Kramers inferred that if we assign the coplanar model to orthohelium and (their version of) the crossed model to parahelium, we can deduce that orthohelium and parahelium states cannot transition among themselves, which was the desired result. In correspondence with Ehrenfest, Bohr spoke of this result as a crucial advantage of his and Kramers's model over Landé's (Mehra and Rechenberg 1982, 407).

However, trouble came in 1923, when Born and Heisenberg embarked on what Heisenberg described to Bohr as "a general investigation of all mechanically allowed orbits of excited helium". As Heisenberg explained, "if in the end the experimentally found terms are not included, then one knows that the mechanics is wrong." (Heisenberg to Bohr 2 Feb. 1923, quoted in Kragh 2012, 255) The results were devastating: in particular, Born and Heisenberg showed that the solution space is such that the principle of mechanical transformability cannot be applied without serious trouble: insofar as stable, multi-periodic solutions are available, they are all mechanically transformable into each other (Born and Heisenberg 1923, 240). And since Born and Heisenberg also show that the only permissible solutions are coplanar and crossed orbits, this meant that the orthohelium and parahelium states of Bohr and Kramers belonged in the same family of orbits after all. As such, they are *realizable in one and the same system*, contrary to what Bohr and Kramers expected. Being realized

[6]One may use the phrase "equivalence class" here as well, but to avoid confusion with the way I have been using this phrase (as indicating a "lumping" of states together), I will use the word "family" in this context.

alongside each other in the same atom means that analogous orbits, such as $2p_o$ and $2p_p$, would be in the same dynamical equivalence class (they would be "lumped together"), which in turn means that they must transition indistinguishably. Therefore, it turned out, Bohr's model could not explain the experimental fact that the orthohelium states never mix with the parahelium states. *This*, it seems, is what brought Bohr's atomic theory to its knees, as it appeared that Bohr's very method of state assignment was doomed. Bohr was the first to declare defeat:

> Born and Heisenberg's investigation may be particularly well suited to provide evidence of the fundamental failure of the laws of mechanics in describing the finer details of the motion of systems with several electrons. (Bohr 1923, 271)[7]

That is, Bohr's orbit-based way of picking out states (as encoded in the Correspondence Principle and the Principle of Mechanical Transformability) lumps together true states that detectably come apart upon transitioning. Thus, the mathematical algorithm through which Bohr's 1918 state assignments are prescribed is *incompatible with the underlying dynamics* of the system. This is a *qualitative* (as opposed to *quantitative*) failure for Bohr's theory, which I have called failure of tracking.

[7]This is not to deny that the Bohr-Sommerfeld theory faced other problems around the same time, not the least of which was the anomalous Zeeman effect (Kragh 2012, 314 ff.).

Appendix C Evidence that Existing Accounts and their Critics Presuppose Referential Transparency

Kitcher (1993, 143-144) seems to make a distinction between those posits that are *genuinely used / deployed* to derive the predictions and those that merely *appear to have been used / deployed*. For instance, the idea is that the luminiferous ether was not genuinely deployed in Fresnel's derivation of his laws and his (novel) prediction of a bright spot in the shadow of a disk (ibid, 145-149). Inspired by Kitcher's language, Lyons (2006; 2009) has coined the term "deployment realism", a view he has fiercely criticized. As the ether example shows, deployment realism goes hand in hand with the notion of retention. Ether was not "deployed" in the derivations and was therefore discarded. Ether *waves* (of the transverse kind) were deployed and were therefore retained. This view therefore implies *retention of entities and/or properties*.[1]

Psillos provides much more explicit and formal criteria for indispensability:

> Suppose that H together with another set of hypotheses H' (and some auxiliaries A) entail a prediction P. H indispensably contributes to the generation of P if H' and A alone cannot yield P and no other available hypothesis H* which is consistent with H' and A can replace H without loss in the relevant derivation of P. (Psillos 1999, 110)

Psillos imposes some "natural epistemic constraints" on H' such as being "independently motivated, non ad hoc, potentially explanatory, etc." (ibid). Since Psillos seems to assume that the hypothesis H will survive intact, his account also requires retention, namely *retention of hypotheses* rather than entities or properties.

Saatsi (2005) appeals to the notion of multiple realizability to make a distinction between "higher-level" and "lower-level" properties, and argues for a realist attitude towards the former, not the latter. Saatsi applies this idea primarily to the case of Fresnel's theory

[1] While this is how Kitcher is understood in most of the literature, I believe this is an incorrect reading of Kitcher's account, likely due to the fact that Ch. 4 of *The Advancement of Science* has not been read as much as its Ch. 5. The only exception to this near-universal misinterpretation that I am aware of is McLeish (2005; 2006) and following him Ladyman (2011, 96). In any case, Kithcer offers an account of *token reference* in which some *token* uses of, say, "dephlogisticated air" referred to oxygen while other tokens of the same phrase referred to nothing. But since according to Kitcher none of the term-*types* of a successful theory (phlogiston, caloric, ether, humor, miasma, etc.) need to refer to anything in reality, this account appears to leave one as ignorant as ever of what to be realist about.

of light. Having presented a "minimal derivation" of Fresnel's laws, Saatsi argues that all Fresnel ever needed for his derivations was to define a quantity we now recognize as the amplitude of the wave and a quantity we now recognize as energy. Thus, according to Saatsi, it is not only ether that was inessential: that light is a wave, that the former quantity is its amplitude and the latter its energy – all these assumptions are eliminable from the derivation according to Saatsi, but the two properties that matter survive. Saatsi's account therefore relies on *retention of properties*.

Drawing inspiration from Chakravartty (2007; 2008), Egg (2016) also suggests that one be realist about properties rather than entities, and that essential content be identified with what Chakravartty calls "detection properties" and idle posits with his "auxiliary properties". Egg argues that detection properties refer to genuine properties of the system, and are therefore expected to show up in future theories in recognizable form, i.e. he assumes *retention of properties*.

As the last of the synchronic accounts, let us briefly review Vickers's account, which is perhaps the most sophisticated of all dispensability criteria. To begin, Vickers (2013) distinguishes between *derivation-external posits (DEPs)* and *derivation-internal posits (DIPs)*. The former are those assumptions that merely *inspire* scientists to advance a certain model or suggest to them a direction to take, but that do not actually *entail* the prediction in question. The latter, on the other hand, are those that are directly (typically deductively) involved in the derivation of the prediction. Vickers's idea is that the working posits are, not the DIPs themselves, but *contained* within the DIPs: "Sometimes only part of such a historically active DIP actually contributes to a given derivational step." (ibid, 201) How is one to extract these "essentially contributing parts" of the DIPs? The basic idea is straightforward: if a prediction has been derived from a proposition P which is a DIP, but it can in fact be derived from a weaker proposition Q which is a logical consequence of P, then credit for the confirmation of the prediction should only go to Q, not P.

Vickers's account allows for some reformulation of the theory, so it is not as transparent as the previous views. Nevertheless, Vickers seems to be assuming that each prediction comes with a fairly unique derivation, presumably the one historically used to derive the prediction.[2]

[2]Peters (2012, 177 ff.) argues there is often no such thing as *the* derivation of a prediction, as a prediction

Vickers does not recommend *reconstructing* this derivation, but merely to logically tighten it up. Once one finds the weaker hypotheses "contained" within the unnecessarily strong premises that are responsible for the predictive success, one should expect these weaker hypotheses to survive future theory change. Thus, Vickers, too, works with a notion of *retention of hypotheses*.

Moving on to diachronic accounts, Harker (2010; 2013) argues that novel predictive successes are really only noteworthy when they signal *progress* over previous theories. Thus, Harker suggests that the essential content can only be determined by comparing a theory with its predecessors and successors and looking for *those posits that resulted in incremental advancements* over the superseded theory. Although Harker admits that this form of realism is more concerned with what is getting closer to the truth than what is already approximately true, he still insists that his thesis "predicts that where past theories improved on the theories they replaced, the parts responsible for such progress will be retained across subsequent cases of theory change." (Harker 2010, 201). Thus, Harker's diachronic account adheres to *entity and hypothesis retention*.

Cordero (2011a; 2011b) similarly argues that during the heyday of a scientific paradigm, there are often too many entrenched metaphysical and methodological prejudices at work to allow for a correct parsing of the theory into working and idle parts. Rather, Cordero argues, it is the wind of subsequent scrutiny that separates the weed from the chaff, although this process might take centuries (2011b, 1128). More recently, Cordero (2013; 2015; 2017) has articulated five "naturalistic" strategies that scientists themselves use to diachronically identify the essential content, which include, for example, hostile probing of a theory's central tenets by its opponents, revision of auxiliary assumptions by its proponents to test the robustness of predictions, and external grounding of theoretical assumptions to give them additional credibility. Despite emphasizing the complexity of this process of theory distillation, Cordero follows the rest of the literature in characterizing the essential content as those components that "we can expect to persist in successor theories" (2011a, 31). As such, Cordero's account is also centered around *retention of entities and hypotheses*.

may be arrived at through several lines of reasoning, thus creating ambiguity in the application of Vickers's method.

Peters (2012) argues that the best account of empirical success is the "unification view", i.e. that success amounts to unifying otherwise disparate bodies of knowledge, and therefore that working posits are those with genuine unificatory power. Peters's method is to start "from the bottom", at the level of observable predictions, and "build up" by adding any premise that would result in further unification of the predictions. While Peters's view has salient differences with diachronic accounts, it shares with them the idea that the virtues of a posit that bestow the honor of essentiality on it can only be revealed over time, as only time can tell whether a given posit will result in further unification. Peters also follows the rest of the literature in expecting the working (i.e. unificatory) posits to be "preserved" or "retained" in future theorizing (e.g. 14, 41, 45 ff., 107). In particular, Peters recognizes two types of preservation: *co-reference* and *structural correspondence*, which indicate cases where a posit is retained because it represents the same real entity and the same structure, respectively, as the current theory.

This brings us to structural realism. Drawing on the case of Fresnel's ether theory, Worrall (1989) argued that while Fresnel was wrong about the *nature* of light, he was very much right about its *structure*, which Worrall claims is captured in Fresnel's equations. And these equations, the argument goes, have been preserved in the modern theory within Maxwell's equations. Thus, Worrall subscribes to *retention of structure*.

Ladyman (2011) claims the case of phlogiston for the structuralist by arguing that while phlogiston theorists were wrong about the *nature* of phenomena of oxydation and reduction, they were quite right about the *relations among these phenomena*, e.g. that burning and rusting are manifestations of the same phenomenon, and that some of these processes are reversible.[3] Ladyman concludes that "much of phlogiston theory is indeed retained in contemporary chemistry" (ibid, 92). Thus, Ladyman's account is also committed to *retention of structure*.

Votsis (2007; 2011a; 2011b) has tried to flesh out the structuralist view of essential content as the "minimally interpreted mathematical parts", where the latter is a form of Ramsey-sentence for the theory. Votsis characterizes the following as a common tenet of

[3]See also Schurz (2009a; 2009b) for a fascinating analysis of the parallels between phlogiston phenomena and electronegativity. Schurz, not a structuralist but a proponent of referential success, suggests that "phlogiston" be taken to refer to "excess electron".

all forms of structural realism: "Approximately all and only operative structural elements of scientific theories have been (and will be) preserved through theory change" (2011b, 11). Once again, we see the commitment to *retention of structure*.

As for the critics, Lyons (2006; 2009) attacks the general notions of essentiality (i.e. that the posit in question was genuinely responsible for the prediction) and indispensability (i.e. that without the posit in question the prediction could not have been derived). Lyons argues that if those are our criteria, then "credit will have to be attributed to all responsible constituents, including mere heuristics (such as mystical beliefs), weak analogies, mistaken calculations, logically invalid reasoning, etc." (2006, 543) To support his claim, Lyons draws on the case of Kepler's predictions regarding planetary motion, specifically his famous second law. Among the assumptions that Kepler deployed in his argument is that the speed of a planet is inversely related to its distance from the sun, i.e. $v \propto \frac{1}{r}$. We know from classical mechanics that the correct formula is $v \propto \frac{1}{\sqrt{r}}$. Kepler's assumption is therefore simply not true – not approximately, not partially, and not in any of its logically weaker consequences. It is totally and hopelessly false. Or take the way Kepler calculated the areas. Seeing as he did not have calculus at his disposal, he attempted to use Archimedean methods and completely botched it (see Lyons 2006, 548). Again, Kepler's math is flat-out wrong; it is not true even approximately or partially. Yet, somehow, Kepler arrived at his correct area law, which was the foundation of Newton's inverse-square formula for gravitation. Kepler, it would seem, arrived at a true and novel prediction via false premises and invalid reasoning. Another example discussed by Lyons (2002) is that of Laplace and Haüy's prediction that all gases expand at the same rate based on several false premises. Lyons's examples are especially powerful in that one is even hard-pressed to claim, a la Vickers (see §3.4.1 above), that any of the premises *contains* truth in one of its logically weaker implications. Note, however, how Lyons assumes that Kepler's and Laplace and Haüy's derivations do not need to be reformulated for us to find their essential content, suggesting his adherence to *retention of hypotheses*.

Stanford makes several inter-related arguments against selective realism, but the core of his complaint seems to be the suspicion that any prospective account of selectivism is bound to require some amount of retrospective cheating, i.e. that prospective criteria cannot

uniquely and informatively pick out the working posits by themselves; the benefit of hindsight is necessary (2006; 2009, 384-385). Trying to determine what is dispensable by examining the theory directly and imagining what can and cannot be removed from it, the idea goes, is just a doomed exercise. The selectivist will end up taking uninformative and often unintelligible positions, such as that there can be vibrations without something that vibrates. I argued above that this happens because one should typically not expect to simply take one part of a coherent theory, rip it apart from its home ontology, and implant it in an imagined future theory without generating nonsense. If such a move is allowed, Stanford argues, then anything is fair game. That is, if the ontological butchering required for posit retention is not even constrained by intelligibility, such accounts of essential content would be too uninformative to discriminate between what can and cannot be eliminated: the selectivist's "filter" for idle posits is too weak and that for working posits too strong (Harker 2010, 196, fn. 18). Note how Stanford assumes that all selectivist accounts must presuppose *retention of entities*.

Bibliography

[1] Abraham, Ralph and Marsden, Jerrold E. (1987), *Foundations of Mechanics*, New York: Perseus Publishing

[2] Aitchison, Ian J. R., David A. MacManus, and Thomas M. Snyder (2004), "Understanding Heisenberg's 'magical' paper of July 1925: A new look at the calculational details", *American Journal of Physics* 72: pp. 1370-1379

[3] Alai, Mario (2014), "Novel Predictions and the No Miracle Argument", *Erkenntnis* 79 (2), pp. 297-326

[4] Badino, Massimiliano (2016), "How to Make Selective Realism More Selective (and More Realist Too)", accessed at: http://philsci-archive.pitt.edu/12523/

[5] Bain, Jonathan (2013), "Category-Theoretic Structure and Radical Ontic Structural Realism", *Synthese* 190 (9), pp. 1621-1635

[6] Bain, Jonathan ([manuscript]), "Towards Structural Realism", accessed at: http://faculty.poly.edu/ jbain/papers/SR.pdf

[7] Barrett, Thomas W. (2014), "On the Structure of Classical Mechanics", *British Journal for the Philosophy of Science* 66, pp. 801–828

[8] Barrett, Thomas W. (2015), "On the Structure of Classical Mechanics", *The British Journal for the Philosophy of Science*, 66, pp. 801-828

[9] Barrett, Thomas W. (2017), "Equivalent and Inequivalent Formulations of Classical Mechanics", accessed at: http://philsci-archive.pitt.edu/13092/

[10] Barrett, Thomas W. and Halvorson, Hans (2016a), "Glymour and Quine on Theoretical Equivalence", *Journal of Philosophical Logic* 45(5), pp. 467–483

[11] Barrett, Thomas W. and Halvorson, Hans (2016b), "Morita Equivalence", *The Review of Symbolic Logic* 9(3), pp. 556-582

[12] Beller, Mara (1983), "Matrix Theory before Schrodinger: Philosophy, Problems, Consequences", *Isis* 74(4): pp. 469-491

[13] Beller, Mara (1999), *Quantum Dialogue: The Making of a Revolution*, Chicago: University of Chicago Press

[14] Beth, Evert Willem (1960), "Semantics of Physical Theories", *Synthese*, 12, pp. 172-175

[15] Bohr, Niels (1913), "On the Constitution of Atoms and Molecules", *Philosophical Magazine Series 6* 26 (151)

[16] Bohr, Niels (1918/2005), *On the quantum theory of line-spectra, Parts I–III*. Dover Publications, Inc.: Mineola, New York

[17] Bohr, Niels (1920), "Über die Serienspektra der Elemente" *Zeitschrift für Physik* 2 (5)

[18] Bohr, Niels (1921), "Atomic Structure", *Nature* 108, pp. 208–209

[19] Bohr, Niels (1923), "Linienspektren und Atombau", *Annalen der Physik* 71, pp. 228-288

[20] Bohr, Niels (1924), "On the Application of the Quantum Theory to Atomic Structure. Part I: The Fundamental Postulates", *Proceedings of the Cambridge Philosophical Society (Supplement)* 22: pp. 1-44

[21] Bohr, Niels, Hendrik A. Kramers, and J. C. Slater (1924), "The Quantum Theory of Radiation", *Philosophical Magazine* 47: pp. 785-802

[22] Born, Max (1922), "Über das Modell der Wasserstoffmolekel", Naturwissenschaften 10, pp. 677-678

[23] Born, Max (1924), "Über Quantenmechanik", *Z. Phys.* 26: p 379

[24] Born, Max (1925), *Vorlesungen über Atommechanik*, Springer–Verlag: Berlin

[25] Born, Max (1955), "Die statistische Deutung der Quantenmechanik" (Nobel lecture delivered on 12/11/1954 at Stockholm), in *Les Prix Nobel en 1954*, Stockholm: Nobel Foundation, pp. 79-90

[26] Born, Max and Heisenberg, Werner (1923), "Die Elektronenbahnen im angeregten Heliumatom", *Zeitschrift für Physik* Band 16, pp. 229-243

[27] Born, Max, and Pascal Jordan (1925), "Zur Quantentheorie aperiodischer Vorgaenge" *Z. Phys.* 33(1)

[28] Boyce, K. (2018), "The Coincidentalist Reply to the No Miracles Argument", *Erkenntnis* 83, pp. 929–946

[29] Boyd, Richard N. (1983), "On the Current Status of the Issue of Scientific Realism", *Erkenntnis* 19(1), pp. 45–90

[30] Boyd, Richard N. (2002) "Scientific Realism", in E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*, available at https://plato.stanford.edu/archives/sum2010/entries/scientific-realism

[31] Brown, Kevin (Manuscript), "Newtonian Precession of Mercury's Perihelion", accessed at: http://www.mathpages.com/home/kmath280/kmath280.htm

[32] Butterfield, Jeremy (2006), "On Symmetry and Conserved Quantities in Classical Mechanics", in *Festschrift for Jeffrey Bub*, eds. W. Demopoulos and I. Pitowsky, Kluwer: University of Western Ontario Series in Philosophy of Science

[33] Camilleri, K. (2009) *Heisenberg and the Interpretation of Quantum mechanics*, Cambridge: Cambridge University Press

[34] Carnap, Rudolph (1966), *Philosophical Foundations of Physics*, ed. M. Gardner, Basic Books: New York

[35] Carnap, Rudolph (1953), "Testability and Meaning", in *Readings in the Philosophy of Science*, eds. H. Feigl and M. Brodbeck, New York: Appleton-Century-Crofts

[36] Cartwright, Nancy (1983), *How the Laws of Physics Lie*, Clarendon, Oxford, UK

[37] Cartwright, Nancy (1999), *The Dappled World: A Study of the Boundaries of Science*, Cambridge University Press: Cambridge, UK

[38] Cassidy, David C. (1979), "Heisenberg's First Core Model of the Atom: The Formation of a Professional Style", *Historical Studies in the Physical Sciences* 10: pp. 187-224

[39] Chakravartty, Anjan (2007), *A Metaphysics for Scientific Realism: Knowing the Unobservable*, Cambridge University Press: New York

[40] Chakravartty, Anjan (2008), "What You Don't Know Can't Hurt You: Realism and the Unconceived", *Philosophical Studies* 137, pp. 149-58

[41] Chakravartty, Anjan (2017), "Scientific Realism", in E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*, available at https://plato.stanford.edu/archives/sum2017/entries/scientific-realism

[42] Chang, Hasok (2003), "Preservative Realism and Its Discontents: Revisiting Caloric", *Philosophy of Science* 70, pp. 902-912

[43] Coffey, Kevin (2014), "Theoretical Equivalence as Interpretative Equivalence", *British Journal for the Philosophy of Science* 65, pp. 821-844

[44] Cordero, Alberto (2011a), "Rejected Posits, Realism, and the History of Science", in H.W. de Regt et al. (eds.), *EPSA Philosophy of Science: 2009*, pp. 23-32

[45] Cordero, Alberto (2011b), "Scientific Realism and the Divide et Impera Strategy: The Ether Saga Revisited", *Philosophy of Science* 78, pp. 1120-1130

[46] Cordero, Alberto (2013), "Theory-Parts for Scientific Realists", in V. Karakostas and D. Dieks (eds.) *EPSA Perspectives and Foundational Problems in Philosophy of Science: 2011*, pp. 153-165

[47] Cordero, Alberto (2015), "On Scientific Realism and Naturalism", *Journal of Philosophical Research* 40, pp. 31-43

[48] Cordero, Alberto (2017), "Retention, Truth-Content and Selective Realism", in E. Agazzi (ed.), *Varieties of Scientific Realism: Objectivity and Truth in Science*, Springer, pp. 245-256

[49] Crepeau, J. (2009), "A Brief History of the T4 Radiation Law", *Proceedings of the ASME 2009 Heat Transfer Summer Conference collocated with the InterPACK09 and 3rd Energy Sustainability Conferences. Volume 1: Heat Transfer in Energy Systems; Thermophysical Properties; Heat Transfer Equipment; Heat Transfer in Electronic Equipment*, San Francisco, pp. 59–65, available at https://doi.org/10.1115/HT2009-88060

[50] Curiel, Erik (2014), "Classical Mechanics is Lagrangian; it is not Hamiltonian", *British Journal for the Philosophy of Science* 65, pp. 269-321

[51] Darrigol, Olivier (1992), *From c-Numbers to q-Numbers: The Classical Analogy in the History of Quantum Theory*, University of California Press: Berkeley, accessed at: http://ark.cdlib.org/ark:/13030/ft4t1nb2gv/

[52] Dawid, Richard (2018), "The No Miracles Argument without Scientific Realism", available at http://philsci-archive.pitt.edu/14559/

[53] Dretske, Fredrich I. (1970), "Epistemic Operators", *Journal of Philosophy* 67, pp. 1007–1023

[54] Dretske, Fredrich I. (1971), "Conclusive Reasons", *Australasian Journal of Philosophy* 49, pp. 1–22

[55] Dretske, Fredrich I. (1980), *Knowledge and the Flow of Information*, MIT/Bradford: Cambridge, MA

[56] Egg, Matthias (2016), "Expanding Our Grasp: Causal Knowledge and the Problem of Unconceived Alternatives", *British Journal for the Philosophy of Science* 67, 115-141

[57] Fine, Arthur (1986), "Unnatural Attitudes: Realist and Instrumentalist Attachments to Science", *Mind* 95, pp. 149-179

[58] Finn, Michael J. (2008), *Classical Mechanics*, Infinity Science Press LLC, Hingham, MA

[59] Franck, James and Einsporn, Erich (1920), "Über die Anregungspotentiale des Quecksilberdampfes", *Zeitschrift für Physik* 2, pp. 18-29

[60] French, Steven (1999), "Models and Mathematics in Physics: The Role of Group Theory", in J. Butterfield and C. Pagonis (eds.), *From Physics to Philosophy*, Cambridge University Press: Cambridge, pp. 187-207

[61] French, Steven (2006), "Structure as a Weapon of the Realist", *Proceedings of the Aristotelian Society* 106, pp. 1-19

[62] French, Steven (2011), "Shifting to Structures in Physics and Biology: A Prophylactic for Promiscuous Realism", *Studies in History and Philosophy of Science, Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 42 (2), pp. 164-173

[63] French, Steven (2011), "Metaphysical Underdetermination: Why Worry?", *Synthese* 180, pp. 205–221

[64] French, Steven and Krause, Décio (2006), *Identity in Physics: A Historical, Philosophical and Formal Analysis*, Oxford University Press: Oxford

[65] French, Steven and Ladyman, James (2003), "Remodelling Structural Realism: Quantum Physics and the Metaphysics of Structure", *Synthese* 136, pp. 31-56

[66] French, Steven and Ladyman, James (2011), "In Defence of Ontic Structural Realism", in A. Bokulich and P. Bokulich (eds.), *Scientific Structuralism (Boston Studies in the Philosophy of Science: Volume 281)*, Springer, pp. 25-42

[67] Friedman, Michael (2001), *Dynamics of Reason*, Center for the Study of Language and Information, Stanford, CA

[68] Gearhart, Clayton A. (2017), "James Franck, the Ionization Potential of Helium, and the Experimental Discovery of Metastable States", *Studies in History and Philosophy of Modern Physics* 60, pp. 95-109

[69] Ghins, Michel (2014), "Bohr's Theory of the Hydrogen Atom: A Selective Realist Interpretation", in L. Magnani (ed.), *Model-Based Reasoning in Science and Technology: Theoretical and Cognitive Issues* (Studies in Applied Philosophy, Epistemology and Rational Ethics Book 8), Springer, pp. 415-429

[70] Giere, Ronald (1988), *Explaining Science: A Cognitive Approach*, The University of Chicago Press, Chicago, IL

[71] Giere, Ronald (1999), *Science without Laws (Science and Its Conceptual Foundations)*, The University of Chicago Press, Chicago, IL

[72] Giere, Ronald (2010), "An Agent-based Conception of Models And Scientific Representation", *Synthese*, 172, pp. 269-281

[73] Giunti, Marco (2006), "Emulation, Reduction, and Emergence in Dynamical Systems", accessed at http://philsci-archive.pitt.edu/2682/

[74] Giunti, Marco (2016), "A Real World Semantics for Deterministic Dynamical Systems with Finitely Many Components", in eds. Laura Felline, Antonio Ledda, Francesco Paoli, Emanuele Rossanese, *New Directions in Logic and the Philosophy of Science*, College Publications, Pittsburgh, PA, pp. 97-110

[75] Glymour, Clark (1970), "Theoretical Realism and Theoretical Equivalence", in *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*

[76] Glymour, Clark (2013), "Theoretical Equivalence and the Semantic View of Theories", *Philosophy of Science* 80(2), pp. 286-297

[77] Goldstein, Herbert (1980 [1950]), *Classical Mechanics*, Addison-Wesley Publishing Company, Reading, MA

[78] Gregory, R. Douglas (2006), *Classical Mechanics: An Undergraduate Text*, Cambridge: Cambridge University Press

[79] Hall, Asaph (1894), "A Suggestion in the Theory of Mercury", *Astronomical Journal*, 319 (XIV), pp. 49-51

[80] Halvorson, Hans (2012), "What Scientific Theories Could Not Be", *Philosophy of Science*, 79 (2), pp. 183-206

[81] Halvorson, Hans (2013), "The Semantic View, if Plausible, is Syntactic", *Philosophy of Science*, 80 (3), pp. 475-478

[82] Halvorson, Hans (2016), "Scientific Theories", accessed at: http://philsci-archive.pitt.edu/11347/

[83] Hamill, Patrick (2014), *A Student's Guide to Lagrangians and Hamiltonians*, New York: Cambridge University Press

[84] Harker, David (2010), "Two Arguments for Scientific Realism Unified", *Studies in History and Philosophy of Science* 41, pp. 192-202

[85] Harker, David (2013), "How to Split a Theory: Defending Selective Realism and Convergence without Proximity", *British Journal for the Philosophy of Science* 64, pp. 79-106

[86] Healey, Richard (2017), *The Quantum Revolution in Philosophy*, Oxford University Press, Oxford

[87] Heisenberg, Werner (1925a), "Über eine Anwendung des Korrespondenzprinzips auf die Frage nach der Polarisation des Fluoreszenzlichtes", *Z. Phys.* 31, pp. 617-626

[88] Heisenberg, Werner (1925b), "Zur Quantentheorie der Multiplettstruktur und der anomalen Zeemaneffekte", *Z. Phys.* 32, pp. 841-860

[89] Heisenberg, Werner (1925c), "Über die quantentheoretische Umdeutung kinematischer und mechanischer Beziehungen", *Z. Phys.* 33: pp. 879-893

[90] Hempel, Carl G. (1958), "The Theoretician's Dilemma: A Study in the Logic of Theory Construction", *Minnesota Studies in the Philosophy of Science*, 2, pp. 173-226

[91] Hempel, Carl G. (1966), *Philosophy of Natural Science*, Englewood-Cliffs, N.J.: Prentice-Hall

[92] Hendry, John (1984), *The Creation of Quantum Mechanics and the Bohr-Pauli Dialogue*, Boston: D. Reidel Publishing Company

[93] Hertz, Heinrich (1894/1956), *The Principles of Mechanics*, trans. D. E.Jones and J. T. Walley, Dover: New York

[94] Hesse, Mary B. (1965), *Forces and Fields*, Totowa, N.J.: Littlefield, Adams & Co.

[95] Holland, P., and Brown, H. R. (2003), "The Non-Relativistic Limits of the Maxwell and Dirac Equations: the Role of Galilean and Gauge Invariance", *Studies In His-*

tory and Philosophy of Science Part B: Studies In History and Philosophy of Modern Physics 34 (2), pp. 161–187

[96] Holton, Gerald (1984), *Themata zur Ideengeschichte der Physik*, Wiesbaden: Springer Fachmedien Wiesbaden

[97] Hoyt, Frank C. (1923), "The Relative Intensity of X-Ray Lines", *Philosophical Magazine* 46: pp. 135-145

[98] Hoyt, Frank C. (1924), "Relative Probabilities of the Transitions Involved in the Blamer Lines of Hydrogen", *Philosophical Magazine* 47: pp. 826-831

[99] Hoyt, Frank C. (1925), "Application of the Correspondence Principle to Relative Intensities in Series Spectra", *Physical Review* 26: pp. 749-760

[100] Jitrik, Oliverio and Bunge, Carlos F. (2004), "Transition Probabilities for Hydrogen-Like Atoms", Journal of Physical Chemistry Reference Data 33 (4), pp. 1059-1070

[101] Kaplan, David M. (2015), "Moving Parts: The Natural Alliance between Dynamical and Mechanistic Modelling Approaches", *Biology and Philosophy*, 30, pp. 757–86

[102] Kaveh, Shahin (2014), "The Incongruent Correspondence: Seven Non-classical Years of Old Quantum Theory", *Studies in History and Philosophy of Modern Physics*, 46 (B), pp. 239-246

[103] Kitcher, Philip (1993), *The Advancement of Science: Science Without Legend, Objectivity without Illusions*, Oxford University Press, New York

[104] Kragh, Helge (1985), "The Fine Structure of Hydrogen and the Gross Structure of the Physics Community", *Historical Studies in the Physical Sciences* 15 (2), pp. 67-125

[105] Kragh, Helge (2012), *Niels Bohr and the Quantum Atom: The Bohr Model of Atomic Structure 1913-1925*, Oxford University Press: Oxford

[106] Kramers, Hendrik A. (1919), *Intensities of Spectral Lines*, Kobenhavn: A. F. Host & Son

[107] Kramers, Hendrik A. (1923), "Über das Modell des Heliumatoms", *Zeitschrift für Physik* 13, pp. 312-341

[108] Kramers, Hendrik A. (1924a), "The Quantum Theory of Dispersion", *Nature* 114, p. 310

[109] Kramers, Hendrik A. (1924b), "The Law of Dispersion and Bohr's Theory of Spectra", *Nature* 133: p. 673

[110] Kramers, Hendrik A., and Werner Heisenberg (1925), "Über die Streuung von Strahlen durch Atome" *Z. Phys.* 31: p. 681

[111] Kuhn, Thomas S. (1963), Interview of Werner Heisenberg by Thomas S. Kuhn on 1963 February 13, Niels Bohr Library & Archives, American Institute of Physics, College Park, MD USA, accessed at www.aip.org/history-programs/niels-bohr-library/oral-histories/4661-4

[112] Lacki, J. (2002) "Observability, Anschaulichkeit and Abstraction: A Journey into Werner Heisenberg's Science and Philosophy", *Fortschritte der Physik* 50(5–7): pp. 440–458

[113] Ladyman, James (1998), "What is Structural Realism?", *Studies in History and Philosophy of Science Part A* 29 (3), pp. 409-424

[114] Ladyman, James (2004), "Modality and Constructive Empiricism: A Reply to van Fraassen", *The British Journal for the Philosophy of Science* 55, pp. 755-765

[115] Ladyman, James (2007), "On the Identity and Diversity of Objects in a Structure", *Proceedings of the Aristotelian Society, Supplementary Volumes* Vol. 81, pp. 23-43

[116] Ladyman, James (2011), "Structural Realism Versus Standard Scientific Realism: the Case of Phlogiston and Dephlogisticated Air", *Synthese* 180, pp. 87-101

[117] Ladyman, James and Ross, Don (2007), *Every Thing Must Go: Metaphysics Naturalized*, Oxford University Press: Oxford (with David Spurrett and John Collier as contributors)

[118] Lakatos, Imre (1978), *The Methodology of Scientific Research Programmes*, eds. John Worrall and Gregory Currie, Cambridge University Press, Cambridge, UK

[119] Landau, Lev D. and Lifshitz, Evgeny M. (2000 [1960]), *Mechanics (Third Edition): Volume I of Course of Theoretical Physics*, Trans. J. B. Sykes and J. S. Bell, Oxford: Reed Educational and Professional Publishing Ltd

[120] Landry, Elaine (1999), "Category Theory: The Language of Mathematics", *Philosophy of Science* 66 (Proceedings), pp. S14-S27

[121] Lange, Marc (2009), *Laws and Lawmakers*, Oxford University Press: New York

[122] Laplace, Pierre Simon (1798-1827), *Mécanique céleste*, trans. Nathaniel Bowditch, *A Treatise of Celestial Mechanics*, Hillard, Gray, Little, and Wilkins, Boston, MA

[123] Laudan, Larry (1981), "A Confutation of Convergent Realism", *Philosophy of Science* 48 (1), pp. 19-49

[124] Laudan, Larry (1984), "Explaining the Success of Science", in J. T. Cushing, C. F. Delaney, G. M. Gutting (eds.), *Science and Reality: Recent Work in the Philosophy of Science* 79, University of Notre Dame Press: Notre Dame

[125] Leconte, Gauvain (2017), "Predictive Success, Partial Truth and Duhemian Realism", *Synthese* 194, pp. 3245-3265

[126] Leplin, Jarrett (1987), "Surrealism", *Mind* XCVI(384), pp. 519–524

[127] Leplin, Jarrett (1997), *A Novel Defense of Scientific Realism*, Oxford University Press: Oxford

[128] Liston, M. N. (2016), "Scientific Realism and Antirealism", in J. Fieser, and B. Dowden (eds.), *Internet Encyclopedia of Philosophy*, available at https://www.iep.utm.edu/sci-real/

[129] Lloyd, Elisabeth (1988 [1994]), *The Structure and Confirmation of Evolutionary Theory*, Princeton University Press, Princeton, NJ

[130] Lloyd, Elisabeth (2013), "Structure of Evolutionary Theory", in *International Encyclopedia of Social and Behavioral Sciences*, eds. W. Durham, 2nd edition, Elsevier, Amsterdam, ND

[131] Lyons, Timothy D. (2002), "Scientific Realism and the Pessimistic Meta-Modus Tollens", in S. Clarke and T. D. Lyons (eds.), *Recent Themes in the Philosophy of Science*, Kluwer: Alphen aan den Rijn, Netherlands, pp. 63-90

[132] Lyons, Timothy D. (2003), "Explaining the Success of a Scientific Theory", *Philosophy of Science* 70, pp. 891-901

[133] Lyons, Timothy D. (2006), "Scientific Realism and the Stratagema de Divide et Impera", *British Journal for the Philosophy of Science* 57, 537-560

[134] Lyons, Timothy D. (2009), "Criteria for Atrributing Predictive Responsibility in the Scientific Realism Debate: Deployment, Essentiality, Belief, Retention...", *Human Affairs* 19, pp. 138-152

[135] Lyons, Timothy D. (2016), "Structural Realism Versus Deployment Realism: A Comparative Evaluation", *Studies in History and Philosophy of Science* 59, 95-105

[136] Maxwell, Grover (1970), "Theories, Perception and Structural Realism", in R. Colodny (ed.), *Nature and Function of Scientific Theories*, University of Pittsburgh Press: Pittsburgh, pp. 3-34

[137] McCain, K., and Kampourakis, K. (2019), *What is Scientific Knowledge?: An Introduction to Contemporary Epistemology of Science*, Routledge: New York

[138] McLeish, Christina (2005), "Scientific Realism Bit by Bit: Part I. Kitcher on Reference", *Studies in History and Philosophy of Science* 36, pp. 667-685

[139] McLeish, Christina (2006), "Realism Bit by Bit: Part II. Disjunctive Partial Reference", *Studies in History and Philosophy of Science* 37, pp. 171-190

[140] McSweeney, Michaela M. (2017), "An Epistemic Account of Metaphysical Equivalence", *Philosophical Perspectives* 30(1), pp. 270-293

[141] Mehra, Jagdish and Rechenberg, Helmut (1982), *The Historical Development of Quantum Theory*, Springer-Verlag: New York

[142] Meyer, Russell (2018), "The Non-mechanistic Option: Defending Dynamical Explanations", *British Journal for the Philosophy of Science*, accessed at https://doi.org/10.1093/bjps/axy034

[143] Mitchell, Sandra (2000), "Dimensions of Scientific Law", *Philosophy of Science* 67: pp. 242–265

[144] Morgan, Mary S. and Margaret Morrison (1999), *Models as Mediators*, Cambridge University Press, Cambridge, UK

[145] Morin, David (2007), *Introduction to Classical Mechanics*, Cambridge: Cambridge University Press

[146] Musgrave, Alan (1988), "The Ultimate Argument for Scientific Realism", in R. Nola (ed.), *Relativism and Realism in Science*, Kluwer Academic Publishers: Dordrecht, pp. 229–252

[147] Musgrave, Alan (2007), "The 'Miracle Argument' For Scientific Realism", *The Rutherford Journal*, **2**, available at http://www.rutherfordjournal.org/article020108.html

[148] Nagel, Ernst (1961), *The Structure of Science*, New York: Harcourt, Brace, & World

[149] Nanni, Luca (2015), "The Hydrogen Atom: a Review on the Birth of Modern Quantum Mechanics", accessed at: https://arxiv.org/ftp/arxiv/papers/1501/1501.05894.pdf

[150] Nasser, Ibraheem (2012), "The Fine Structure of the Hydrogen Atom", accessed at: http://faculty.kfupm.edu.sa/PHYS/imnasser/Phys551/spin_orbit.pdf

[151] Navarro, Luis and Perez, Enric (2006), "Paul Ehrenfest: The Genesis of the Adiabatic Hypothesis, 1911–1914", *Archive for History of Exact Sciences* 60, pp. 209-267

[152] Newton, Isaac (1687 [1999]), *The Principia: Mathematical Principles of Natural Philosophy*, University of California Press, Berkeley, CA

[153] Nguyen, James (2017), "Scientific Representation and Theoretical Equivalence", *Philosophy of Science* 84(5), pp. 982-995

[154] Niiniluoto, Ilkka (1987), *Truthlikeness*, Reidel: Dordrecht

[155] Niiniluoto, Ilkka (1998), "Survey Article: Verisimilitude: The Third Period", *The British Journal for the Philosophy of Science* 49: pp. 1–29

[156] Niiniluoto, Ilkka (1999), *Critical Scientific Realism*, Oxford University Press: Oxford

[157] Niiniluoto, Ilkka (2003), "Content and likeness definitions of verisimilitude", in J. Hintikka, T. Czarnecki, K. Kijania-Placek, T. Placek, and A. Rojszczak (eds.), *Philosophy and Logic: In Search of the Polish Tradition*, Kluwer: Dordrecht, pp. 27–36

[158] Niiniluoto, Ilkka (2018), *Truth-seeking by Abduction*, Springer Nature Switzerland AG: Cham, Switzerland

[159] NIST Helium Tables, *Basic Atomic Spectroscopic Data, Helium. National Institute of Science and Technology*, accessed at: http://physics.nist.gov/PhysRefData/Handbook/Tables/heliumtable2.htm

[160] North, Jill (2009), "The 'Structure' of Physics: A Case Study", *Journal of Philosophy*, 106 (2), pp. 57-88

[161] Norton, John D. (2000), "How we know about electrons", in R. Nola and H. Sankey (eds.), *After Popper, Kuhn and Feyerabend*, Kluwer: Alphen aan den Rijn, Netherlands, pp. 67-97

[162] Norton, John D. (2002 [2010]), "A Survey of Inductive Generalization", accessed at: https://www.pitt.edu/ jdnorton/papers/Survey_ind_gen.pdf

[163] Norton, John D. (2003), "Causation as Folk Science", *Philosophers' Imprint*, 3 (4), pp. 1-22

[164] Norton, John D. (2008), "The Dome: An Unexpectedly Simple Failure of Determinism", *Philosophy of Science*, 75, pp. 786-98

[165] Norton, John D. (2016), *The Material Theory of Induction*, accessed at: https://www.pitt.edu/ jdnorton/papers/material_theory/material.html

[166] Norton, John D. and Bain, Jonathan, "What Should Philosophers of Science Learn from the History of the Electron?", in A. Warwick (ed.), *Histories of the Electron: The Birth of Microphysics*, pp. 451-465

[167] Nozick, Robert (1981), *Philosophical Explanations*, Harvard University Press: Cambridge, MA

[168] Onishi, Yukinori (2017), "Defending the Selective Confirmation Strategy", *Studies in History and Philosophy of Science* 64, pp. 1-10

[169] OpenStax (2016), *Chemistry*, OpenStax CNX, accessed at: http://cnx.org/contents/85abf193-2bd2-4908-8563-90b8a7ac8df6@9.311

[170] Pais, Abraham (1991), *Niels Bohr's Times*, Oxford University Press: Oxford

[171] Park, Seungbae (2001), *Scientific Realism vs. Scientific Antirealism*, The University of Arizona, PhD dissertation

[172] Park, Seungbae (2003), "Ontological Order in Scientific Explanation", *Philosophical Papers* 32(2), pp. 157–170

[173] Pashby, Thomas (2012), "Dirac's Prediction of the Positron: A Case Study for the Current Realism Debate", *Perspectives on Science* 20 (4), pp. 440-475

[174] Pauli, Wolfgang (1919), "Merkurperihelbewegung und Strahlenablenkung in Weyls Gravitationtheorie" *PGV* 21: pp. 742-750

[175] Pérez, Enric (2009), "Ehrenfest's Adiabatic Theory and the Old Quantum Theory, 1916–1918", *Archive for History of Exact Sciences* 63, pp. 81-125

[176] Pérez, Enric and Valls, Blai Pié (2015), "Ehrenfest's Adiabatic Hypothesis in Bohr's Quantum Theory", accessed at: https://arxiv.org/ftp/arxiv/papers/1502/1502.03022.pdf

[177] Peters, Dean (2012), "How to Be a Scientific Realist (If at All): A Study of Partial Realism", PhD diss., London School of Economics

[178] Peters, Dean (2014), "What Elements of Successful Scientific Theories Are the Correct Targets for 'Selective' Scientific Realism?", *Philosophy of Science* 81, pp. 377-397

[179] Psillos, Stathis (1999), *Scientific Realism: How Science Tracks Truth*, Routledge, New York, NY

[180] Psillos, Stathis (2001), "Predictive Similarity and the Success of Science: A Reply to Stanford", *Philosophy of Science* 68, pp. 346–355

[181] Putnam, Hilary (1963a/1975), "'Degree of Confirmation' and Inductive Logic", in P. A. Schilpp (ed.), *The Philosophy of Rudolf Carnap*, LaSalle, Ill.: Open Court, pp. 761–784. Reprinted in *Mathematics, Matter and Method* (1975), pp. 270–292

[182] Putnam, Hilary (1963b/1975), "Probability and Confirmation", *The Voice of America Forum Lectures, Philosophy of Science Series* 10, United States Information Agency: Washington, D.C.. Reprinted in *Mathematics, Matter and Method* [1975], pp. 293–304

[183] Putnam, Hilary (1975), *Mathematics, Matter, and Method*, Cambridge University Press: New York

[184] Putnam, Hilary (1978), *Meaning and the Moral Sciences*, Routledge Kegan & Paul: Boston

[185] Putnam, Hilary (1981), *Reason, Truth and History*, Cambridge University Press: New York

[186] Quine, Willard V. O. (1975), "On empirically equivalent systems of the world", Erkenntnis 9(3), pp. 313-328

[187] Rohlf, James William (1994), *Modern Physics from α to Z^0*, Wiley: Hoboken, NJ

[188] Rosaler, Joshua (2015a), "Local Reduction in Physics", *Studies in History and Philosophy of Modern Physics*, 50, pp. 54-69

[189] Rosaler, Joshua (2015b), "Interpretation Neutrality in the Classical Domain of Quantum Theory", *Studies in History and Philosophy of Modern Physics*, 53, pp. 54-72

[190] Ross, Lauren (2015), "Dynamical Models and Explanation in Neuroscience", *Philosophy of Science*, 82, pp. 32-54

[191] Roush, Sherrilyn (2005), *Tracking Truth: Knowledge, Evidence, and Science*, Oxford University Press: New York

[192] Ruetsche, Laura (2002), "Interpreting Quantum Theories", in P. Machamer and M. Silberstein (eds.), *The Blackwell Guide to the Philosophy of Science*, Blackwell: Malden, pp. 199-226

[193] Ruetsche, Laura (2015), "The Shaky Game +25, or on locavoracity", *Synthese*, 192, pp. 3425–3442

[194] Ruetsche, Laura (2016), "Physics and Method", in eds. Herman Cappelen, Tamar Szabó Gendler, and John Hawthorne, *The Oxford Handbook of Philosophical Methodology*, Oxford University Press, Oxford, UK

[195] Saatsi, Juha (2005), "Reconsidering the Fresnel–Maxwell Theory Shift: How the Realist Can Have Her Cake and EAT It Too", *Studies in History and Philosophy of Science* 36, pp. 509-538

[196] Saatsi, Juha (2009), "Scientific Realism and Historical Evidence", in H.W. de Regt et al. (eds.), *EPSA Philosophy of Science: 2009*, pp. 329-340

[197] Saatsi, Juha (2017), "Replacing Recipe Realism", *Synthese* 194, pp. 3233-3244

[198] Sakellariou, Vassilis (2011), "Restoring Continuity in Theory Change: The Kepler-to-Newton Case", *Journal of General Philosophy of Science* 42, pp. 109-127

[199] Sankey, Howard (2001), "Scientific Realism: An Elaboration and a Defense", *Theoria* 98, pp. 35-54

[200] Schaffner, Kenneth F. (1969), "Correspondence Rules", *Philosophy of Science* 36(3), pp. 280-290

[201] Schurz, Gerhard (2009a), "When Empirical Success Implies Theoretical Reference: A Structural Correspondence Theorem", British Journal for the Philosophy of Science 60, pp. 101-133

[202] Schurz, Gerhard (2009b), "Preservation of Empirical Success and Intertheoretical Correspondence: Justifying Realism without the No Miracles Argument", in A. Hieke, H. Leitgeb (eds.), *Publications of the Austrian Ludwig Wittgenstein Society - New Series Vol. 11: Reduction – Abstraction – Analysis*, Ontos Verlag: Heusenstamm

[203] Smart, J. J. C. (1963), *Philosophy and Scientific Realism*, Routledge & Kegan Paul: London

[204] Smart, J. J. C. (1968), *Between Science and Philosophy*, Random House: New York

[205] Smith, Sheldon R. (2008), "Symmetries and the explanation of conservation laws in the light of the inverse problem in Lagrangian mechanics", *Studies in History and Philosophy of Modern Physics* 39: pp. 325-345

[206] Sneed, Joseph D. (1971 [1979]), *The Logical Structure of Mathematical Physics*, Reidel, Dordrecht

[207] Sommerfeld, Arnold (1923), "The Model of the Neutral Helium Atom", *Journal of the Optical Society of America* 7 (7), pp. 509-515

[208] Sommerfeld, Arnold (1924), "Grundlagen der Quantentheorie und des Bohrschen Atommodelles", *Naturwissenschaften* 12: pp. 1047-1049

[209] Sommerfeld, Arnold, and Werner Heisenberg (1922), "Die Intensität der Mehrfachlinien und ihrer Zeemankomponenten", *Z. Phys.* 11: pp. 131-154

[210] Stanford, Kyle (2000), "An Antirealist Explanation of the Success of Science", *Philosophy of Science* 67 (26)

[211] Stanford, Kyle (2003), "No Refuge for Realism: Selective Confirmation and the History of Science", *Philosophy of Science* 70(5), pp. 913-925

[212] Stanford, Kyle (2006), *Exceeding Our Grasp: Science, History, and the Problem of Unconceived Alternatives*, Oxford University Press, Oxford, UK

[213] Stanford, Kyle (2009), "Grasping at Realist Straws", *Metascience* 18, pp. 355-390 (review symposium with Psillos and Saatsi)

[214] Stanford, Kyle (2017, unpublished), "A Middle Path Forward in the Scientific Realism Debate", in *Science, Value and Pragmatism: A Conference in Honour of Philip Kitcher*, Columbia University

[215] Stanford, Kyle (2018), "A Fond Farewell to 'Approximate Truth"', *Spontaneous Generations: A Journal for the History and Philosophy of Science* 9(1), pp. 78–81

[216] Stegmüller, Wolfgang (1979), *The Structuralist View of Theories*, Springer, Berlin/Heidelberg/New York

[217] Suppe, Frederick (2000), "Understanding Scientific Theories: An Assessment of Developments, 1969-1998", *Philosophy of Science*, 67 (Proceedings), pp. S102-S115

[218] Swanson, Noel, and Hans Halvorson (2012), "On North's 'The Structure of Physics'", accessed at: http://philsci-archive.pitt.edu/9314/

[219] Taylor, John R. (2005), *Classical Mechanics*, Herndon, VA: University Science Books

[220] Teh, Nicholas J. and Tsementzis, Dimitris (2017), "Theoretical Equivalence in Classical Mechanics and Its Relationship to Duality", *Studies in History and Philosophy of Modern Physics* 59, pp. 44–54

[221] Thackray, Arnold (1970), *Atoms and Powers: An Essay on Newtonian Matter-Theory and the Development of Chemistry*, Harvard University Press, Cambridge, MA

[222] Thagard, Paul (2000), *Coherence in Thought and Action*, MIT Press: Cambridge, MA

[223] Thagard, Paul (2007), "Coherence, Truth, and the Development of Scientific Knowledge", *Philosophy of Science* 74, pp. 28-47

[224] Tolman, Richard C. (1925), "Note on the Problem of the Quantitative Formulation of Bohr's Correspondence Principle", *Philosophical Magazine* 49: pp. 130-136

[225] Tulczyjew, Wlodzimierz M. (1974), "Hamiltonian Systems, Lagrangian Systems, and the Legendre Transformation", *Symposia Mathematica* 14, pp. 247–258.

[226] Tulczyjew, Wlodzimierz M. (1977), The Legendre Transformation, Annales de L' I.H.P., Section A(1), pp. 101–114

[227] Uppal, Shveta (ed.) (2006), *Physics: Textbook for Class XI, Part I*, National Council of Educational Research and Training

[228] Van der Waerden, B. L. (1967), *Sources of Quantum Mechanics*, Amsterdam: North-Holland Publishing Company

[229] van Fraassen, Bas C. (1970), "On the Extension of Beth's Semantics of Physical Theories", *Philosophy of Science*, 37 (3), pp. 325-339

[230] van Fraassen, Bas C. (1980), *The Scientific Image*, Oxford University Press, Oxford, UK

[231] van Fraassen, Bas C. (1989), *Laws and Symmetry*, Clarendon, Oxford, UK

[232] van Fraassen, Bas C. (2006), "Structure: Its Shadow and Substance", *The British Journal for the Philosophy of Science* 57 (2), pp. 275-307

[233] Van Vleck, John Hasbrouck (1924), "The Absorption of Radiation by Multiply Periodic Orbits, Part I", *Physical Review* 24, reprinted in G. Holton (ed.), *Sources of quantum mechanics*, 1967, North-Holland Publishing Co.: Amsterdam

[234] Van Vleck, John Hasbrouck (1926), *Quantum Principles and Line Spectra*, Bulletin of the National Research Council, v. 10 pt. 4 no. 54

[235] Vickers, Peter (2012), "Historical magic in old quantum theory?", *European Journal of Philosophy of Science* 2, pp. 1-19

[236] Vickers, Peter (2013), "A Confrontation of Convergent Realism", *Philosophy of Science* 80, pp. 189–211

[237] Vickers, Peter (2016), "Towards a Realistic Success-to-Truth Inference for Scientific Realism", *Synthese*, accessed at: https://doi.org/10.1007/s11229-016-1150-9

[238] Vickers, Peter (2017), "Historical Challenges to Realism", in J. Saatis (ed.), *The Routledge Handbook of Scientific Realism*, Routledge: London, pp. 48-59

[239] Vickers, Peter (2018), "Disarming the Ultimate Historical Challenge to Scientific Realism", *The British Journal for the Philosophy of Science*, accessed at: https://doi.org/10.1093/bjps/axy035

[240] Votsis, Ioannis (2007), "Uninterpreted Equations and the Structure-Nature Distinction", *Philosophical Inquiry* XXIX (1-2), pp. 57-71

[241] Votsis, Ioannis (2011a), "The Prospective Stance in Realism", *Philosophy of Science* 78 (5), pp. 1223-1234

[242] Votsis, Ioannis (2011b), "Structural Realism: Continuity and Its Limits", in A. Bokulich and P. Bokulich (eds.), *Scientific Structuralism (Boston Studies in the Philosophy of Science)*, Springer, pp. 105-118

[243] Votsis, Ioannis and Schurz, Gerhard (2012), "A Frame-Theoretic Analysis of Two Rival Conceptions of Heat", *Studies in History and Philosophy of Science* 43, 105-114

[244] Wald, Robert (1994), *Quantum Field Theory in Curved Spacetime and Blackhole Thermodynamics*, = University of Chicago Press, Chicago and London

[245] Wallace, David (2012), *The Emergent Multiverse: Quantum Theory according to the Everett Interpretation*, Oxford University Press, Oxford, UK

[246] Weatherall, James O. (2016a), "Are Newtonian Gravitation and Geometrized Newtonian Gravitation Theoretically Equivalent?", *Erkenntnis* 81, pp. 1073–1091

[247] Weatherall, James O. (2016b), "Understanding Gauge", *Philosophy of Science* 83, pp. 1039–1049

[248] Wilson, Mark (manuscript), "Background Considerations for Wandering Significance", accessed at: http://www.philosophy.pitt.edu/sites/default/files/

[249] Wigner, Eugene (1979), *Symmetries and Reflections*, Ox Bow Press: Woodbridge, CT

[250] Williams, Porter (2019), "Scientific Realism Made Effective", *British Journal for the Philosophy of Science* 70, pp. 209–237

[251] Wilson, Mark (2008), *Wandering Significance*. Oxford University Press, Oxford, UK

[252] Wilson, Mark (2018), *Physics Avoidance*. Oxford University Press, Oxford, UK

[253] Wimsatt, William C. (2007), *Re-Engineering Philosophy for Limited Beings*, Harvard University Press, Cambridge, MA

[254] Wolff, Johanna (2014), "Heisenberg's Observability Principle", *Studies in History and Philosophy of Modern Physics* 45: pp. 19–26

[255] Woodward, James (2003), *Making Things Happen: A Theory of Causal Explanation*, Oxford University Press: New York

[256] Woodward, James (2013), "Laws, Causes, and Invariance", in S. Mumford and M. Tugby (eds.), *Metaphysics and Science*, Oxford University Press: Oxford, pp. 48–72

[257] Woodward, James (2018), "Laws: an Invariance-Based Account", in W. Ott and L. Patton (eds.), *Laws of Nature*, Oxford University Press: Oxford, pp. 158–180

[258] Woodward, James (2020 [preprint]), "Sketch of Some Themes for a Pragmatic Philosophy of Science", accessed at: http://philsci-archive.pitt.edu/17694/

[259] Woodward, James, and Mark Wilson (2019), "Counterfactuals in the Real World", *Proceedings of Conference on Algorithms and Complexity in Mathematics, Epistemology, and Science*, pp. 269-294

[260] Worrall, John (1989), "Structural realism: The best of both worlds?", *Dialectica* 43 (1-2), pp. 99-124

[261] Worrall, John (1994), "How to Remain (Reasonably) Optimistic: Scientific Realism and the 'Luminiferous Ether"', *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association Volume One: Contributed Papers*, pp. 334-342

[262] Worrall, John (2007) "Miracles and Models: Why Reports of the Death of Structural Realism May Be Exaggerated", *Royal Institute of Philosophy Supplements* 82 (61), pp. 125-154

[263] Wray, Brad K. (2007), "A Selectionist Explanation for the Success and Failures of Science", *Erkenntnis* 67 (1), pp. 81-89

[264] Wray, Brad K. (2010), "Selection and Predictive Success", *Erkenntnis* 72 (3), pp. 365-377

[265] Wüthrich, Adrian (2016), "Heisenberg's Umdeutung: A Case for a (Quantum-)Dialogue Between History and Philosophy of Science", in Tilman Sauer and Raphael Scholl (eds.), *The Philosophy of Historical Case Studies*, Springer International Publishing Switzerland Yoshimi, Jeffrey (2012), "Supervenience, Dynamical Systems Theory, and Non-reductive Physicalism", *The British Journal for the Philosophy of Science*, 63 (2), pp. 373-398

[266] Zednik, Carlos (2011), "The Nature of Dynamical Explanation", *Philosophy of Science*, 78, pp. 238-63