Title
Scientific Understanding and Pragmatic Rationality

Author
Bhakthavatsalam, Sindhuja

Publication Date
2015

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA, SAN DIEGO

Scientific Understanding and Pragmatic Rationality

A dissertation submitted in partial satisfaction of the requirements
for the degree Doctor of Philosophy

in

Philosophy (Science Studies)

by

Sindhuja Bhakthavatsalam

Committee in charge:

Professor Nancy Cartwright, Chair
Professor Craig Callender
Professor Hasok Chang
Professor Cathy Gere
Professor Tal Golan
Professor Christian Wüthrich

2015

The Dissertation of Sindhuja Bhakthavatsalam is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

Chair

University of California, San Diego

2015

Dedication

For Ma, who’s stood by me like a rock;

For Pa and Sumi, who taught me to be philosophical;

For Sundar Sarukkai, who inspired me to pursue philosophy;

For Deepthi, who held my hand and argued with me.

Table of Contents

Signature Page ……………………………………………………………………...iii

Dedication.………………………………………………………………………..iv

Table of Contents ……………………………………………………………………v

Acknowledgements.…………………………………………………………………viii

Vita ……………………………………………………………………………………xi

Abstract of the Dissertation ……………………………………………………….xiv

Part I – Understanding in Science…………………………………………………1

1. Empirical Adequacy and Non-Epistemic Values..…………….……………………...2

1.1. Introduction …………………………………...2

1.2 The Primacy of Empirical Adequacy………………………………………...5

1.3 The Importance of Values other than Empirical Adequacy………………...10

1.4 Values Trumping Empicial Adequacy………………………..................12

1.4.1 Purpose-specific Values and Pluralism………………………………...13

1.4.2 Pragmatic Functionalism………………………………………………18

1.5 Conclusion.……………………………………………………………....21

References…………………………………………………………………22

2. Scientific Understanding – Models, Theories, Stories …………………………...24

2.1 Introduction…………………………………………………………………24

2.2 Truth, Empirical Adequacy, Understanding ……………………………….27

2.3 Conclusion ………………………………………………………………46

References……………………………………………………………………47

3. Ontological Plausibility and Scientific Understanding……………………………..50

3.1 Overview ……………………………………………………………...50

3.2 Ontological Plausibility and Scientific Understanding …………………….52

3.2.1 Background and Exegesis ……………………………………………...52

3.2.2 Reflection and Critique ………………………………………………...62

3.3 Conclusion ……………………………………………………………72

References……………………………………………………………………73

4. Ontological Principles and Epistemic Activities ………………………………...74

4.1 Overview ……………………………………………………………...74

4.2 Ontological Principles and Epistemic Activities……………………………..75

4.2.1 Background and Exegesis ……………………………………………...75

4.2.2 Reflection and Critique ………………………………………………...87

4.3 Conclusion ……………………………………………………………97

References……………………………………………………………………97

Part II – Understanding of Science…………………………………………………99

5. Philosophy of Science in the time of Pierre Duhem…………………………….100

5.1 Introduction ………………………………………………………….100

5.2 Science and Philosophy in the time of Pierre Duhem – an Overview…….102

5.2.1 Historical Backdrop – A Brief Account ………………………………102

5.2.2 Karl Pearson on Science, Knowledge, and Reality …………………...106

5.2.3 Henri Poincaré on Truth in Physics…………………………………..119

5.2.4 On Ernst Mach's Economy and (Anti) Realism in Science…………..126

5.3 Conclusion…………………………………………………………...138

References………………………………………………………………..138

6. Duhem's Philosophy of Physical Theory…………………………..………..140

6.1 Introduction …………………………………………………………….140

6.2 Duhem, Natural Classification, and Structure..…………………...……141

6.2.1 Duhem on Physical Theory – An Overview ………………………….141

6.2.2 Was Duhem concerned with Structure? ………………………………144

6.2.3 Conclusion ……………………………………………………………158

6.3 The Rationale Behind Duhem's Natural Classification.…………………..159

6.3.1 Background …………………………………………………………...159

6.3.2 Introduction to Duhemian Natural Classification …………………….161

6.3.3 Duhem's Rationale: existing views centered on the success of novel predictions ……………………………………………………………….167

6.3.4 The TNC Rationale: An initial response and its distance from the success of novel predictions …………………………………………………..171

6.3.5 The Rationale behind TNC elaborated: making sense of pursuing physical theory …………………………………………………………..174

6.3.6 Hasok Chang and Rationalizing Epistemic Activities.………………..194

6.3.7 Duhem's Natural Ontological Attitude ……………………………….201

6.3.8 Conclusion ……………………………………………………………203

References……………………………………………………….……………..204

Acknowledgements

I'm deeply indebted to Nancy Cartwright for her incredible support and guidance through the years. The time and critical attention that she has invested in this dissertation is remarkable, and learning from her has been an extremely rewarding experience. Through copious written feedback on multiple drafts of these chapters, spending long hours in her office nailing down arguments, and many a delightful conversation while walking across Californian deserts and the English countryside, Nancy has had a profound influence on my work. I'm very grateful for her supervision.

Craig Callender and Chris Wüthrich have both provided very valuable comments on several aspects of this dissertation. Through critical feedback on my writing and many helpful conversations through the years, their guidance goes beyond this dissertation and I believe they have been instrumental in helping me mature as a philosopher. I'm also thankful to Craig for pointing me to some very useful references for my work. Tal Golan helped me critically engage with many historical aspects of the work on Duhem here, that I would have otherwise overlooked. I'm thankful to him for his useful comments, particularly on Chapter 5 of this dissertation. Cathy Gere has been a big support and I thank her for feedback on chapters as well as for patiently listening to me and providing helpful pointers as I brainstormed ideas.

Hasok Chang has been a huge influence and this dissertation would not have been possible without his work. I'm very thankful to him for some long and extremely stimulating discussions on Duhem and pragmatic rationality, as well as for some very involved and detailed comments on several aspects of this dissertation.

My visit to Durham University in 2013 was a very enriching experience. I'm particularly grateful to Ian James Kidd for the keen interest he took in my work, and for the many insightful conversations on Duhem, values in science, virtue epistemology, and epistemic humility. I also had valuable discussions with Peter Vickers. It has been a pleasure discussing some of the ideas in this work with people I met at conferences, including Kyle Stanford, Anjan Chakravartty, and Juha Saatsi.

Faculty members in the philosophy department and the Science Studies program, particularly Eric Watkins and Bob Westman, have helped shape ideas in this dissertation in various ways. I'm also thankful to the philosophy department and Science Studies program for providing excellent support for the duration of my degree.

A slightly modified version of Section 6.3 has been published as the following paper and I thank two anonymous referees for some very helpful comments:

Bhakthavatsalam, Sindhuja. 'The Rationale Behind Pierre Duhem's Natural Classification' – Studies in History and Philosophy of Science (2015) 51 11-21

Section 6.2 is being prepared for publication and to appear as:

Bhakthavatsalam Sindhuja. “Duhem, Natural Classification, and Structure”

Portions of Chapters 1 and 2 are being prepared for publication by Nancy Cartwright and myself and to appear as:

Bhakthavatsalam, Sindhuja; Cartwright, Nancy. ‘What’s so good about empirical adequacy?’

Despite such amazing support from highly distinguished scholars, particularly the dissertation committee, I am aware that this work is not fully satisfactory. I offer my sincere apologies. This dissertation cannot repay the intellectual debt I’ve incurred to those listed above, but it’s a start.

Vita

<u>Areas of Specialization</u>: Philosophy of Science, Scientific Understanding, Science and Values

<u>Areas of Competence</u>: Science Studies, Ethics and Society, Philosophy of Physics

<u>Education</u>

- **Doctor of Philosophy**: Philosophy and Science Studies; University of California, San Diego – Awarded June 2015
- **Master of Science**: Physics; Indian Institute of Technology, Guwahati, India – Awarded May 2007
- **Bachelor of Science**: Physics, Chemistry, Mathematics; Mount Carmel College, Bangalore University, India – Awarded June 2005

<u>Other Affiliations</u>

- Visiting research scholar, Center for the Humanities Engaging Science and Society (CHESS), Durham University, Durham, UK: April – July 2013

<u>Publications</u>

Edited Volumes

- *Unconceived Alternatives and Scientific Realism* – Co-edited with Ian James Kidd (Durham University) for special issue of Synthese – In preparation

Articles

- 'The Rationale Behind Pierre Duhem's Natural Classification' – Studies in History and Philosophy of Science Vol. 51 (June 2015), pp. 11-21
- 'What's so good about empirical adequacy?' – Coauthored with Nancy Cartwright in *Unconceived Alternatives and Scientific Realism* for special issue of Synthese – In preparation

Book Reviews

- *Discovery and classification in Astronomy* by Steven J. Dick – Philosophy of Science Vol. 82 No. 3 (July 2015), pp. 520-525

Presentations

- *Why not to Aim for Truth in Theory* – with Nancy Cartwright (Invited) – Workshop on Non-Alethic Aims of Inquiry, University of St Andrews, Scotland, October 2014
- *Duhem and Structural Realism* – Philosophy Graduate Conference with Anjan Chakravartty, UC San Diego, May 2014
- *The Rationale behind Pierre Duhem's Natural Classification* – Science Studies Colloquium, UC San Diego, April 2014
- *Plausibility, Intelligibility, and Theory Appraisal* – Society for the Social Studies of Science, San Diego, October 2013
- *Plausibility as a Cognitive Value* – Society for Philosophy of Science in Practice, Toronto, Canada, June 2013
- *What's so Good about Empirical Adequacy?*– with Nancy Cartwright (Invited) – Conference on Unconceived Alternatives and Scientific Realism with Kyle Stanford, Durham University, England, July 2013
- *Ontological Plausibility as a Desideratum in Theory Evaluation* – British Society for Philosophy of Science, Exeter, UK, July 2013
- *In Search of Norms for Privatized Science* – UT Dallas Center for Values in Medicine, Science, and Technology, April 2012
- *Duhem and Conditional Plausibility Realism* – Berkeley-Stanford-Davis Philosophy Graduate Conference, April 2012
- *Duhem: A Plausibility Realist* – Science Studies Graduate Conference with Hasok Chang, UC San Diego, May 2011

Awards

- Dissertation Fellowship: Science Studies Program, UC San Diego, Fall 2014, Fall 2013
- Dissertation Fellowship: Department of Philosophy, UC San Diego, Spring 2013; Fall 2012; Spring 2012.

Workshops

- Selected to participate in the Geneva Summer School in the Philosophy of Physics, Geneva, Switzerland, July 2009

Teaching Experience

As sole instructor:

Philosophy Department, UC San Diego

- Introduction to Philosophy – Nature of Reality: Summer 2014

As Teaching Assistant:

Philosophy Department, UC San Diego

- Scientific Reasoning: Spring 2009, Spring 2015
- Ethics and Society: Winter 2015, Spring 2014, Winter 2014, Winter 2009
- Introduction to Philosophy – Nature of Reality: Winter 2013
- Philosophy and the Environment: Summer 2012
- Introduction to Logic: Summer 2010, Fall 2008

Culture, Art, and Technology – Writing Program at Sixth College, UC San Diego

- Public Rhetoric and Practical Communication (Upper Division): Winter 2012, Fall 2011, Spring 2011
- Climate, Technology and Culture: Winter 2011
- Thinking with Things: Fall 2010
- A History of Time: Time and Modern Society: Spring 2010
- Laws of Men and Laws of Nature: Winter 2010
- Invention of the Person: Fall 2009

Professional Service

- Reviewer for Philosophy of Science (2013)
- Reviewer for International Studies in the Philosophy of Science (2015)
- Commentator at the American Philosophical Association Pacific Division Meeting, April 2014
- Member of the American Philosophical Association
- Member of Philosophy of Science Association

ABSTRACT OF THE DISSERTATION

Scientific Understanding and Pragmatic Rationality

by

Sindhuja Bhakthavatsalam

Doctor of Philosophy in Philosophy (Science Studies)

University of California, San Diego, 2015

Professor Nancy Cartwright, Chair

This dissertation is on scientific understanding, pragmatic rationality, and values in science. I argue for an 'egalitarian' picture of theoretic values and aims in science. Both anti-realists and realists demand that a good theory be empirically adequate. In my dissertation I focus on one important job for science that often does not care for empirical adequacy – understanding. I look at some important modes of achieving understanding in science and show that they often have very little to do with empirical adequacy. After looking at understanding got from the 'products' of science *viz.* theories and models, I

focus on understanding in relation to *activities and practices*. This is centered on Hasok Chang's (2009) work on ontological principles and the intelligibility of epistemic activities. Chang's view is that in order for our (pragmatically chosen) epistemic activities to make sense to us, we cannot deny certain corresponding ontological principles, for if we did, the activities would be rendered unintelligible.

Finally I look at Duhem's philosophy of physical theory. I situate Duhem among some of his key historical contemporaries Mach, Pearson, and Poincaré and engage in a comparative analysis of these 20^{th} century historian-scientist-philosophers all of whom are widely perceived as paradigmatic instrumentalists. I then launch into Duhem's philosophy. Duhem believed that the aim of physical theory is to classify experimental laws, and that this classification progressively approaches a *natural*, underlying classification – call this latter the thesis of natural classification. First I argue that contrary to the views of many scholars, Duhem was not a *structural* realist. I contend that Duhem was not concerned with structure as it is generally construed, *viz.* the mathematical form of equations. Duhem was rather concerned with the classification of laws by theory. Finally, I look at Duhem's rationale behind his idea of natural classification. I situate Duhem in Chang's activity-and-principle scheme and argue that for Duhem, in order for the physicist to make sense of her activity of theorizing, she had to affirm the 'principle' or thesis of natural classification. This way I make the move from (Changian) understanding *in* science, to (Duhemian) understanding *of* science

Part I – Understanding in Science

1. Empirical Adequacy and Non-Epistemic Values

1.1 Introduction

If there is one virtue of scientific theories and models that has near-universal appeal, it is empirical adequacy. Empirical adequacy is regarded to be an essential virtue in an acceptable theory by scientific realists and antirealists alike. While realists want to make the case that we can infer the truth of theories from their empirical adequacy, antirealists argue that empirical adequacy is as far as we can get to the truths about our world – or for some others, empirical adequacy is all there is to get. But both camps value telling things as they are. One demands that a good theory tell things as they are about the empirical facts that it implies; the other, about all the facts it implies, both empirical and theoretical[1]. For both camps, other virtues of theories, such as simplicity, elegance etc., are rightly seen as subordinate to empirical adequacy because they are not constitutive of the immediate aim.

But why should we take non truth-indicative virtues to be subordinate to empirical adequacy in any way? There are many equally – or even more – important aims and jobs for the sciences to do than aim at empirical adequacy (and for the realists, ultimately truth). Most realists and antirealists wouldn't agree, but I contend that the burden is on them to say why exactly empirical adequacy is the most important aim. Theories can be

[1] Some other values like internal consistency have also sometimes been thought to be truth indicative. But throughout this chapter (and the next) I will only focus on empirical adequacy

accepted for different reasons in different contexts to do different jobs. Of course, if a job – like description or prediction – requires that a theory be empirically adequate, then it must be; otherwise, it need not be. Ultimately, a theory must be suited to the job at hand. I present arguments from the literature for an egalitarian picture of values in science, following Larry Laudan and more recently Hasok Chang and Matthew Brown. Science has multiple jobs and aims and each of these may demand – and *not* demand, and sometimes even demand the *absence* of – different values including empirical adequacy. I will present arguments for the views that

1. Empirical adequacy alone is an inadequate criterion for evaluating theories and that several non truth-indicative virtues play at least as large a part in science as the truth-indicative empirical adequacy. It is often believed that it is more okay for a theory to not be simple or have wide explanatory power or give us understanding, than it is for it to not be empirically adequate. But a theory can be equally demanded to be simple/ elegant/ explanatory in addition to being empirically adequate – in a specific situation depending on the task at hand, simplicity might be as crucial as empirical adequacy. Theoretical values should suit our ends.
2. Sometimes we might even *not* care about a theory being empirically adequate: other values can in fact trump empirical adequacy.

In Section 1.2 I start with the familiar arguments in support of empirical adequacy as an essential virtue of theories In Section 1.3 I present arguments for the point that other values are required in addition to empirical adequacy, and often these other values play as big a role as does empirical adequacy. Importantly, for those who think these other values are not truth tracking – once we are in the business of choosing one model over another on the basis of some value X (say, simplicity) that is not truth conducive, it looks as if we are happy to settle for false models on the grounds they have a certain value X for us. And if they are false but have X for us, why not also settle for models that are false and not empirically adequate (except where they must be) that nevertheless have X? So finally in Section 1.4 I give arguments for the view that empirical adequacy does not necessarily trump other virtues: there are tasks that science can be entrusted with that don't demand empirical adequacy and where other values trump. I will continue this last line in the next chapter where I present cases of theories/ models in science that are not empirically adequate, but that help with an important task for science: to provide understanding.

In this chapter I present general arguments from the literature against taking empirical adequacy as the most important theoretical virtue. *Much of this is intended to be normative and distinct from the project of giving a descriptive account of science that fits my picture.* I contend that there are no good reasons for a blanket adoption of empirical adequacy as the be-all-and-end-all feature of good scientific practice. In the following chapter I give examples of theories and models that are not empirically adequate by normal standards but serve the important purpose of giving us understanding.

Before I move forward, I should clarify what I mean by empirical adequacy. A theory or model is said to be empirically adequate, I take it, when most of when most of its empirical predictions are borne out, where some distinction between empirical and theoretical phenomena is supposed. But just how many empirical claims that a theory makes have to get correct; and how correct do they have to be, in order for us to say that it is empirically adequate? Where do we draw the line between empirical adequacy and empirical inadequacy? Of course, there is no straight answer. But just because we don't know exactly where to draw the line, it's not that there is no line to be drawn. As we shall see in a number of the cases in this chapter and the next, there are good theories and models (that do various things like give us understanding) that are *clearly* empirically inadequate by any reasonable standard.

1.2 The Primacy of Empirical Adequacy

It's a familiar point that a theory rightly may be accepted or rejected due to its theoretical virtues. But what everyone agrees on, realists and antirealist, is that accepted theories ought to be at least empirically adequate. As a representative example, I present Heather Douglas' (2006) view. Douglas argues that demonstrably false theories cannot be acceptable. She distinguishes between minimal epistemic criteria: "necessary aspects of scientific theories", and "desirable attributes" or ideal desiderata. According to her, minimal epistemic criteria include empirical adequacy, predictive competency and internal consistency:

> **E**: If a theory does not meet the basic requirements of epistemic values (e.g. empirical adequacy, predictive competency, internal consistency), it is doubtful that one has met the basic requirements for any scientific theory. Theories that obviously fail to match with evidence, or contain inconsistencies that allow one to predict anything from the theory, are not a good basis for empirical belief. For this reason, these values are more accurately described as baseline epistemic *criteria* for science than *values*. If a theory does not meet these criteria, it should be rejected or held in abeyance until made adequate. (2006, 6, emphasis as in original)

Douglas goes on to say that values like simplicity and explanatory power, i.e. those that are not *baseline criteria*, cannot be taken as grounds for accepting a theory or hypothesis as well supported or true – they are not epistemic. And according to her, epistemic criteria need to be urgently met. She admits that in practice scientists may aim for characteristics like simplicity and potential explanatory power even in the face of failings in the minimal criteria of empirical adequacy and internal consistency. She also thinks these other non-epistemic values can be useful in theory evaluation: values can guide by helping us decide "sufficiency of warrant" and how to manage *uncertainty.* For instance, take the value of simplicity: according to her, one should not believe the simpler of two theories to be more likely to be true, but simplicity can be considered a 'hedge' against uncertainty: "…any flaws in the theory are more likely to be found sooner rather than later because the theory is easier to work with, and the theory is easier to expand into new areas." (2006, 6)

But, she maintains that entertaining an empirically inadequate theory for other values

> …must be done with the full acknowledgement that the theory is inadequate as it stands, and that it must be corrected to meet the minimum

> requirements as quickly as possible. Although philosophers like to quip that every scientific theory is 'born falsified', no scientist should be happy about it. (2012, 6)

According to Douglas, baseline epistemic criteria such as empirical adequacy are "genuinely truth assuring, in the minimal sense that their absence indicates a clear epistemic problem." (2012, 4)

All this fuss about empirical adequacy seems to indicate that it is somehow less serious of a problem if a theory is not simple/ does not have adequate explanatory power/ does not give understanding than if it is not empirically adequate. But I think that in a particular situation depending on the goal at hand, other values that serve that goal might also be "baseline criteria". As we'll see in the next section, Larry Laudan shows that these other values loom at least as large as empirical adequacy – if not more – in scientific practice. Further, note as above (in quote E), that Douglas says that failure to match with evidence is not a good basis for *empirical belief* (by which I take it she means belief in empirically true claims) and acceptance of the theory as true. But arguably, science is not just about empirical beliefs. Arguably, there are values in science that don't contribute to forming empirical beliefs, but are still valuable – and they can be valued even in the face of failing in truth-indicators[2] like empirical adequacy. We turn to science

[2] It should be noted here that although internal consistency has been for long considered to be a truth-indicative virtue and a "baseline criterion", there have been several cases of scientists doing productive things with self-contradictory theories as Peter Vickers (2013) has shown in his book *Understanding Inconsistent Science*. Scientists have even made useful and accurate predictions with inconsistent theories. For instance, Bohr's model of the atom is widely regarded as internally inconsistent – as da Costa and French (2003) note, the model uses classical mechanics for the dynamics of the stationary electron but

to be able to build a bridge, to understand the structure of the atom, and so on. Not all of these tasks we can entrust science with necessarily involve belief and truth, and not all of them require us to have a theory or model that only and strictly makes empirically true claims. As I show in the next chapter, a theory might only be very modestly empirically adequate and hence may not be very good by way of belief that it is true, but it could still help with certain important tasks. And there is no reason to think that helping us form good/ true beliefs is the most important aim of science. So Douglas' claim that empirical adequacy is a baseline criterion for *science* in general, based on the fact that it is a baseline criterion for empirical belief – which is just one aspect of science – is questionable. I contend that realists and antirealists who deny this claim and defend the primacy of empirical adequacy owe an account of what science is, and an explanation for why they specially pick out empirical adequacy and think it is of special interest.

Douglas' views are quite similar to Bas van Fraassen's (1980) views with respect to empirical adequacy and belief. Van Fraassen allows that one may in pursuit of usable and useful theories demand other 'pragmatic' virtues in a theory, like simplicity, scope and explanatory power. But at the end of the day, no matter what goals you may have as a *scientist*, the crux of van Fraassen's philosophy – constructive empiricism – takes it that the aim of *science* is to provide empirically adequate theories. So for van Fraassen, getting an empirically adequate theory is a fundamental aim of science, regardless of the pragmatic aims of scientists. Van Fraassen, calls empirical adequacy a "semantic" virtue

---

quantum mechanics for the transitions of the electron between discrete states. Yet, it was remarkably empirically adequate.

of a theory: it has to do with the relation between a theory and (parts of) the world. This is again a Douglas-like position in that it takes "pragmatic" aims – i.e. aims other than that of pursuing empirical adequacy – as somewhat lesser aims. Only, while Douglas seems to think that empirical adequacy is truth-indicative, van Fraassen doesn't think that empirical adequacy gives us grounds to believe that a theory is true. Nevertheless, they are united in the view that empirical adequacy tells us what to believe about the world – in Douglas' case it seems to be some underlying truth, and in van Fraassen's case, claims about observables – and is somehow more fundamental than other values like simplicity, explanatory power and scope. But van Fraassen does not really give any good reasons for valuing empirical adequacy so highly.

There are transparent reasons why we could be concerned with a variety of different goals, such as truth itself – for its own sake, for forming true beliefs, – either about empirical phenomena, or some special phenomena that matter to us, or about what theoretical entities there are, or how they behave, and so on. As I had said earlier, if truth and empirical adequacy serve good purposes, then we should most certainly pursue them. Indeed, it is hard to think of a purpose we would want a scientific claim/ hypothesis/ model/ theory to serve that would not require truth of *anything*. For instance, if I want to use a theory to build a bridge, I will probably want its prediction that the bridge will bear a given weight to be true even if the model gets a lot of other things wrong, both at the theoretical and the empirical level But van Fraassen gives us no reasons for why we should be constructive empiricists, nor for why we should be concerned with belief at all. I don't find any convincing arguments that forming (empirically) true beliefs about the world is intrinsically of high worth: it could very well be the goal of many scientists, but I don't see good reasons for mandating that it be the goal of everyone in the business of science. It can and should be pursued when needed, but not otherwise.

1.3 The importance of values other than empirical adequacy

It is now a common view that values other than empirical adequacy play a big role in scientific practice alongside empirical adequacy. But there is dissent about just how big a role they can play. Arguably, these other values can be just as important as empirical adequacy and need not be subordinate to it. In this section I present an argument for the view that values other than empirical adequacy play at least as big a role as empirical adequacy. In the next section I present arguments for the view that other values can play *bigger* roles than empirical adequacy – they can be favored *over* empirical adequacy.

Larry Laudan (2004) points out that values other than empirical adequacy have a big presence in science. One of Laudan's central points is that scientists don't always reject a theory because it is refuted and don't always support a theory because it is true or empirically adequate. Theories are widely valued for their explanatory power, simplicity etc. He gives several examples from scientific practice that support this point:

> Steady-state cosmology was rejected in the 1960s not because it had been refuted but because it offered no account of the cosmic background radiation discovered at Bell Labs. The uniformitarian theories of Hutton,

> Playfair, and Lyell were rejected by most nineteenth century geologists, not because they faced massive refutations, but because they steadfastly refused to say anything about how the earth might have evolved from its primitive initial condition to the condition of habitability. (2004, 17)

Importantly for Laudan, explanatory power (among other values) has nothing to do with truth and is hence not epistemic. He underlines that "scientists have expectations about good theories that go well beyond worries about their veracity" (2004, 19) According to Laudan, values like explanatory power (as he pictures explanation) that are not truth-indicative factors "loom at least as large" in theory evaluation as ones like empirical adequacy that are truth-indicative. He adds that a theory does not have to be false to be bad: "A theory may be bad because it fails the test of possessing the relevant nonepistemic virtues." (2004, 19) So his central point I take it, is that there isn't a hierarchy of theoretical values with empirical adequacy sitting at the top: rather, other, non truth-indicative values play *equal* roles in theory evaluation.

The main takeaway from Laudan I think, is that empirical adequacy (and perhaps internal consistency) need not be the only "minimal criteria" in Douglas' terms: other values like explanatory power can also be as minimal a criterion depending on the situation and the aim/ goal/ task at hand – these other values can also, depending on the situation, be nonnegotiable – like empirical adequacy is taken to be. Does this mean that Douglas' scheme of minimal criteria and ideal desiderata simply doesn't work? Since I'm saying that values like simplicity can also be minimal criteria in specific situations given the task at hand, is there really no distinction to be drawn between minimal criteria and ideal desiderata? I believe there is – but it's just that Douglas' criterion of minimal criteria being epistemic or truth-indicative doesn't work. I propose that something be termed a minimal criterion if it is *always* required to be fulfilled. Surely, empirical adequacy does not *always* have to be met by theories and models as we'll see in the next section and in the next chapter. (But there is one value that I think should belong to minimal criteria, and that's intelligibility – that will be the topic of Chapters 3 and 4.)

Laudan is, however, non-committal on one key issue: that of empirical *in*adequacy and falsity of theories. While he says that scientists may in fact regard truth to be an important value, and contends that "what cannot be gainsaid is that there are other virtues of theories that loom at least as large in theory evaluation as truth does." (19), he doesn't say what one should do with a theory that is false. Could it still be accepted/ entertained for its other values like explanatory power? Or should it be rejected? i.e., it is not clear if Laudan wants of a theory, *both* epistemic and non-epistemic values, or if he would (at least in some situations) accept a theory for its non-epistemic values even if it is shown to be (empirically) false.

1.4 Values trumping empirical adequacy

In the last section I discussed the case for the view that values other than empirical adequacy can play as big a role in scientific practice. Here I discuss arguments for the view that other values can in fact trump empirical adequacy. A theory can get just very few empirical facts in its domain right, and hence – by normal standards – empirically inadequate, but can still be a good theory owing to other values.

1.4.1 Purpose-specific values and Pluralism

My overall view should be clear by now. I believe that theoretical virtues – and the role they play – should be pragmatically chosen, based on the task at hand. This implies that a theory can be valued for virtues other than empirical adequacy, and importantly, it can be valued even when it is not empirically adequate by normal standards. As Chang (2014) persuasively argues, values should suit the task at hand and the effectiveness of the value should be measured by our success at the activity – and this claim I think should be fairly uncontroversial. According to this view then, preferring empirical adequacy over all else assumes an overly narrow construal of success – perhaps in terms of prediction. But success can be constituted in several different ways to be determined by different epistemic communities. For instance, a grossly empirically inadequate model of the atom can be used to explain atomic structure to a five year old. Is success at the task here to be understood in the same way that it is understood for a practicing physicist studying atomic structure? Of course not. One might argue that we're not really "doing science" in the first situation. But why not? I contend that we're as much engaging in science while explaining a scientific idea to a five year old, as we are when we study the idea as a practicing scientist. It's simply that the goals, measures of success, and therefore values, are different in each case.

But there are situations where empirical adequacy is not valued as much as it is usually thought to be, even among practicing scientists. A good example of such a situation is the story of the triumph of Lavoisier's oxygen theory over the preceding phlogiston theory, discussed by Chang (2014). Chang notes, contrary to the popular view that Lavoisier's theory superseded the phlogiston theory on empirical grounds, that "Lavoisier's system was very successful in attaining elegance, unifying power and explanatory power, while it was lacking in empirical adequacy as it had many anomalies" (2014, 230) He brings to attention the facts that

> Lavoisier confidently predicted in vain that muriatic acid (hydro-chloric acid, HCl, in modern terms) would be decomposed into oxygen and the "muriatic radical"; two other non-existent radicals, fluoric and boracic, can be seen in Lavoisier's table of simple substances… Lavoisierian responses to similar anomalies of prussic acid (HCN) and sulphuretted hydrogen (H_2S) not containing any oxygen also had no progressive outcomes. And in neutralizing Berthollet's challenge about the combustion of gunpowder, Lavoisier again only managed *ad hoc* hypotheses unaccompanied by successful novel predictions. Lavoisierians also made pretty un-progressive responses to the discovery that not only oxygen but also chlorine gas supported combustion, but no other known gases did. (2014, 54)

By usual standards, I think the above failures would discredit Lavoisier's theory as being quite seriously empirically inadequate – but the theory has enjoyed a privileged status, and this is not without good reasons. As above, the theory had several other virtues that worked in its favor, for they were deemed important by members of the epistemic communities engaged in combustion research at the time.

But was it chosen despite being *less* empirically adequate *than the phlogiston theory*? Perhaps despite its empirical shortcomings, it was more empirically adequate than its predecessor, or at the least, just as empirically adequate? If so, then it cannot be claimed that other values trumped empirical adequacy in the choice of the oxygen theory over the phlogiston theory. But as Chang points out, deciding if one theory is more or less empirically adequate then another is itself not a straightforward process: "Theories are not simply "tested against evidence"; we must always choose *where* they ought to be tested against evidence—which is to say, where we most wish them to be empirically successful." (2014, 20) Further:

> there were different standards according to which one or the other was better supported by empirical evidence. In a way, this is only an indication that evidential support is not a straightforward matter of logical or probabilistic connections between theory and observation, but a complex rela- tionship mediated by epistemic values, which can be divergent and contextual. (2014, 29)

It's also worth nothing that Chang quotes Alan Musgrave as claiming that "While Lavoisier was failing, Priestley was having great success with the 1766 version of phlogistonism. . . . the most impressive experiment of all came in early 1783." (quoted in Chang, 2009, 11) (Chang tells us that "the most impressive experiment" referred to the confirmation of the phlogistonist prediction that calxes would be reduced to metals by heating in inflammable air.)

Chang (2009, 2014) presents an interesting, fundamental scheme of how in general we go about engaging in "epistemic[3] activities". The idea is that an epistemic activity is paired with a corresponding ontological principle that renders the activity

---

[3] By "epistemic activity" Chang does not mean an activity that is in any way related to truth. So "epistemic" is probably not an appropriate term – Chang could have used a more general term like "cognitive".

meaningful. This scheme is entirely pragmatic: the idea simply is that regardless of whether the principles are really true or empirically grounded, we need to presume them for *our* (respective) *pragmatically chosen* pursuits to make sense to us. I discuss this at length in Chapter 4, but I will give a few examples of this idea here to illustrate the central point of this chapter. Chang (2009) discusses the very simple case of what he calls 'testing by overdetermination': to test a theory by comparing a theoretical value of a physical quantity with an experimental value, we need to presume what he calls 'the principle of single value'. The idea is that to *intelligibly* engage in the activity of such testing, we cannot deny that a physical quantity, say temperature, can have only one value at a given place and time. For instance, if the theory predicts the temperature at a certain point in space and time to be 5^{o} C and the experimental result is 10^{o} C, our activity of testing would not be intelligible if we accepted the possibility that temperature could take both values. For the sake of intelligibility, there has to be the presumption that one of the values is wrong/ unacceptable. The crux is that this principle of single value is not empirically grounded (and nor is it based on requirements of logic). Chang points out that it is not the case that we go around taking measurements of various quantities and check if we get a single value each time. But the principle still seems necessary to us – and this necessity arises, Chang contends, from our pragmatic requirements of engaging in the activity of testing-by-overdetermination. It can be argued that one of our most successful theories ever – quantum mechanics – incorporates violations of the principle of single value, and is still generally accepted as one of our best scientific theories because of its amazing empirical adequacy. But this doesn't take away the fact that the principle of single value *has* to be presumed *for the purpose of* testing-by-overdetermination. There are several such examples of activity-principle pairs that Chang offers and I discuss them in Chapter 4. Here I will leave you with just one other example.

In detailing the Lavoisier episode in the history of combustion science, Chang (2014) notes that in contrast to the earlier phlogiston theory, Lavoisier put forward weight-based arguments about the combustion of substances. He showed that out of 100g of water we get 85g of oxygen and 15g of hydrogen, and further, when we combine 85 g of oxygen and 15g of hydrogen, we get 100g of water. This was considered to be clear proof of the idea that water is made of oxygen and hydrogen, as opposed to the phlogiston story, for the latter gave no good explanation for "why phlogistication should make water less dense than dephlogistication does, or a precise measure of how much phlogiston went into the part of water that became hydrogen (or came out of the part that became oxygen)." (2014, 36) Chang then points out that Lavoisier's reasoning was based on two very important assumptions: that weight is a good measure of the amount of all chemical substances, and that weight is conserved. However, Chang notes that neither of these principles was exactly well founded: there were exceptions to the first within Lavoisier's own framework (since light and caloric, the first two in his list of simple substances didn't have weight), and the second principle is a violation of $E=mc^2$. Chang also cites William Nicholson as having in fact doubted the degree of accuracy of Lavoisier's weight measurements. (2014, 36) And of course, Lavoisier's oxygen-hydrogen ratio – 85:15 – is nowhere close to the currently accepted ratio of 8:1. But – Chang underlines – Lavoisier and his followers *presumed* the above two principles about weight for the principles were required for their system of experimental practices. Here then is a case of a celebrated revolution in science, the very building blocks of which are not empirically grounded.

One key lesson that Chang wants to drive home is pluralism: theories/ models/ claims/ principles can be used or discarded based on our pragmatic needs, and different – even incompatible – theories or principles can be used for different purposes. A theory that is empirically adequate may be used for the purpose of prediction, but a 'rival' theory that is not empirically adequate may be used for some other purpose that doesn't require empirical adequacy – like explanation or understanding.

1.4.2 Pragmatic Functionalism

Let's step back for a bit and ask again, why exactly is empirical adequacy so highly valued? A common answer is that if we let other values trump, we run the risk of slipping into wishful thinking; but if we let evidence trump, we take the objective route without coloring our judgment with our interests and biases. Matthew Brown (2012) challenges just this view and argues against the view that evidence – and hence values or criteria that are based on it, like empirical adequacy – are lexically prior[4] to other values.

[4] I understand "lexical priority" along the lines of John Rawls' (1999) well-known use of the term in his *A Theory of Justice,* to refer to a method for ordering principles in a way that is comparable to alphabetic ordering. It is an ordering "which requires us to satisfy the first principle in the ordering before we can move on to the second, and the second before we consider the third, and so on. A principle does not come into play until those previous to it are either fully met or do not apply. A serial ordering avoids, then, having to balance principles at all; those earlier in the ordering have absolute weight, so to speak, with respect to later ones, and hold without exception" (1999, 38)

He starts by delineating the “gap” argument from underdetermination and the “error” argument from inductive risk as the starting points for the introduction of values in science. According to the former, there is an inevitable gap between theory and evidence: evidence underdetermines theory as famously argued by Pierre Duhem. Hence, say the value proponents, this gap needs to be filled with values. According to the latter, when there is an uncertainty associated with a theory given the evidence, values help decide “sufficiency of warrant” as Douglas puts it: One is likely to find out if something is wrong with a theory sooner rather than later if the theory exhibits values such as simplicity or broad scope.

Brown points out, “Both arguments begin from a situation where the evidence is fixed and take values to play a role in the space that is left over”. (2012, 7) He observes that value proponents adhere to this strict priority of evidence over values probably due to the carefulness to avoid wishful thinking in science: there is a worry that “inquirers might rig the game in favor of their preferred values”. (2012, 8) He then argues that the lexical priority of evidence over values is unfounded and that this worry is misguided for two reasons. First:

> … evidence can turn out to be bad in all sorts of ways: unreliable, unrepresentative, noisy, laden with unsuitable concepts and interpretations, or irrelevant for the question at hand; the experimental apparatus could even have a cord loose. More importantly, we may be totally unaware of why the evidence is bad; [As critics] of strict falsificationism and empiricism have shown, we already have reason to adopt a more egalitarian account of the process of testing and certification, independent of the question about the role of values. We might get off to a

> better start if we thought about how to fit values into this sort of picture of testing. (2012, 9)

Second, value judgments are generally not the same as merely “valuing”. While the latter may be a mere expression of preference, the former can be adopted for good reasons:

> Just as the good (partly empirical) reasons for adopting a theory, hypothesis, or background assumption can give us good reasons to reinterpret, reject, or maybe even ignore evidence apparently in conflict with them (under certain conditions), so too with a good value judgment. If evidence and values pull in opposite directions on the acceptance of a hypothesis, then we should not always be forced to follow the (putative) evidence.” (2012, 10)

Further, he points out following Elizabeth Anderson that “value judgments say something like “try it, you’ll like it”— a testable hypothesis” (2012, 10) An epistemic community can adopt a theoretical virtue and see how it fares for the purpose of a chosen task – this way a theoretical virtue can in fact be tested for its usefulness. (How it fares is of course to be decided by the community again – on pragmatic grounds.)

Brown underlines that we can avoid the problem of wishful thinking without resorting to the lexical priority of evidence if we are not dogmatic about our values (and nor should we be dogmatic about evidence, of course). He suggests that we could adopt a ‘pragmatist functionalism about inquiry’, which “differentiates the functional roles of evidence, theory, and values in inquiry.” (2012, 11) On this account, all three have to be coordinated in their functional roles in a way that solves the problem at hand, and each is revisable in the face of new experience. This is very much in line with Chang’s view of non-empirical ontological principles coordinated with pragmatically chosen epistemic activities.

For Brown, values and evidence are treated as "mutually necessary, functionally differentiated, and rationally revisable components of certification." (2012, 11) Importantly, such an account would "allow that *evidence may be rejected* because of lack of fit with a favored hypothesis and compelling value-judgments, but only so long as one is still able to effectively solve the problem that spurred the inquiry." (2012, 11, emphasis mine)

Note that Brown's approach is rooted in pragmatism: the problem at hand to be solved dictates which values to aim for, and how big a role evidence and values each play. Chang too suggests something similar – pragmatic pluralism: we could entertain multiple conflicting values all at once, each one for a different *purpose.* While intelligibility may help in understanding, simplicity may help solve an equation and empirical adequacy may help with predictions and belief-formation.

1.5 Conclusion

In conclusion, there are compelling arguments for not treating empirical adequacy as the holy grail of evaluation of scientific claims, theories, models. There seem to be no objective, non-arbitrary reasons for touting empirical adequacy to be the ultimate goal for the sciences. Other values may play an equal, or even bigger role in many situations. Both pragmatist functionalism and pluralism are valuable approaches in this regard.

In the next chapter, I turn to scientific understanding and the role of empirical adequacy and truth. I argue that understanding is an aim of science that often does not demand truth and empirically adequacy, while demanding other values like visualizability, simplicity, and various kinds of explanation. (But there can be accounts of understanding that do require truth and/or empirical adequacy – in Chapter 3 I argue that Chang's plausibility account of understanding is one such.)

Portions of Chapters 1 and 2 are being prepared for publication by Nancy Cartwright and myself and to appear as:

Bhakthavatsalam, Sindhuja; Cartwright, Nancy. 'What's so good about empirical adequacy?'

References

- Brown, Matthew J. (2012) 'Values in Science Beyond Underdetermination and Inductive Risk' Philosophy of Science 80 829– 839
- Chang, Hasok (2014) *Is Water H2O?: Evidence, Realism and Pluralism* Boston Studies in the Philosophy and History of Science (Book 293) Springer
- Chang, Hasok (2001) 'How to take Realism beyond Foot-Stamping' *Philosophy* Vol 76 No. 295 Cambridge University Press 5–30

- Chang, Hasok (2009) 'Ontological Principles and the Intelligibility of Epistemic Activities', in Henk de Regt, Sabina Leonelli, and Kai Eigner, eds., *Scientific Understanding: Philosophical Perspectives* University of Pittsburgh Press 64–82

- Da Costa, Newton C.A. and French, Steven (2003) *Science and Partial Truth* Oxford University Press

- Douglas, Heather (2006) 'Norms for Values in Scientific Belief Acceptance' Philosophy of Science Assoc. 20th Biennial Mtg PSA Contributed Papers

- Douglas, Heather (2012) 'The Value of Cognitive Values' Philosophy of Science Assoc. 23rd Biennial Mtg PSA Contributed Papers

- Laudan, Larry (2004) 'The Epistemic, the Cognitive and the Social' Machamer, P. and Wolters, G eds., *Science, Values, and Objectivity* University of Pittsburgh Press Pittsburgh, and Universitatsverlag, Konstanz 14-23

- Rawls, John (1999) *A Theory of Justice* (Revised Ed.) Harvard University Press

- Van Frassen, Bas (1980) *The Scientific Image* Oxford University Press

- Vickers, Peter (2013) *Understanding Inconsistent Science* Oxford University Press

2. Scientific Understanding: Models, Theories, Stories

2.1 Introduction

How do we understand the world through science? How do scientific artifacts such as models, explanations, and theories give us understanding of natural phenomena? What features make these artifacts good vehicles of understanding? Many philosophers have held that although these may be perfectly good questions, understanding does not fall within the ambit of philosophy of science. That is to say, it has not been considered a job of philosophy of science to look into how scientists comprehend and make sense of the world via explanations, models, and theories. Until recently, understanding, taken to consist in the celebrated 'aha ' experience, was considered to be purely pragmatic, subjective and phenomenological. Hempel (1965) and more recently Trout (2007) have argued this. And while explanatory and predictive features of theories and models have studied extensively, not much has been said about what makes them good tools for understanding. But from Newton's dissatisfaction with action-at-a-distance – for it didn't give him a good understanding of gravitation – to constructing models for explaining and visualizing, it is undeniable that understanding has always been a key concern in scientific practice[5]. Moreover Michael Friedman (1974) made the influential argument that although understanding may be psychological, that does not make it subjective. He argues that just as the concept of rational belief is, unlike truth, psychological, yet also considered objective, understanding can also be both psychological and objective – in the

[5] See de Regt and Dieks (2005) for an argument for this.

sense that “what is scientifically comprehensible is constant for a relatively large class of people.” (1974, 8) So, Friedman urged, scientific understanding deserves attention by philosophers of science.

Understanding, as Henk de Regt, Sabina Leonelli, and Kai Eigner (2009) have pointed out, is a three-term relation involving a model, explanation, or other vehicle, the target object or phenomenon that we want to understand, and an agent, the understander. Although I believe agency, the human aspect, is a very important part of studying understanding, there seems to be a rough division of labor. Epistemology focuses on agency – what characteristics an agent must have and what they must do in order to understand (cf. the recent spate of work in epistemology on ‘grasping’ an explanation as central to understanding). Philosophy of science has by contrast not been much concerned with what is in the agent’s head, or what they are doing in understanding something but rather with the public products of science that can provide understanding, the vehicles of understanding, which might, in a very broad sense be labeled explanatory devices’ or ‘explanations’. Both kinds of concern are important but they require different tools and concepts and build on large disparate bodies of previous work. In keeping with the setting of this dissertation in the philosophy of science and in science studies, my discussion of understanding will focus on the vehicles of understanding – especially on the models and theories that provide understanding – and their features, and the relationship they bear with the respective target objects or phenomena. Specifically, I look at the relevance of truth and empirical adequacy to understanding. Through many examples from scientific practice, I show that models and explanations need not be true, and in some cases even empirically adequate, to be good vehicles of understanding. That is, we can gain understanding of (parts of) the world from models and theories that are not descriptively true or empirically adequate.

There is a venerable tradition of demanding that science provide understanding of the world. Here I won't suppose a single exhaustive account of understanding, for I don't think it's possible to have one. There have been such views though, like Lord Kelvin famously declared, "It seems to me that the test of `Do we or not understand a particular subject in physics?' is, `Can we make a mechanical model of it?'" But such views/ accounts are specific precisifications of the concept made in specific sciences at specific times. Scientists pursue understanding in different ways and of different kinds in different contexts. I'll look at a few examples of models and theories from scientific practice that give us different kinds of understanding, and see how truth and empirical adequacy fare with them. In this process, I will look at some currently available accounts of understanding in the literature, which capture reasonable, though different, senses of understanding. It would be wonderful to pre-define understanding and then see if and/or how each of these models/ theories gives us understanding in that pre-defined sense. But as above, I don't think understanding can be captured in a single definition. But I promise that even if *prima facie* it is not clear how the model/ theory gives understanding, it will be arguable that it does – and I give the arguments in each case.

2.2 Truth, Empirical Adequacy, and Understanding

I'll look at models and theories that are/ can be false and empirically inadequate to different degrees, and the different ways in which we gain understanding in each case. I classify these different kinds of models/ theories and the understanding we get from each into four categories. But first I'll look at the case of understanding afforded by idealized models that make several false assumptions but are nevertheless quite empirically *adequate*: despite the false assumptions most of the empirical predictions of these models are borne out. I'll look at two accounts of how such a model gives us understanding: the first is that it gives us understanding by helping us predict and manipulate. The second is a more fundamental account of how it gives us understanding: the idea is that *it gets right the main causes* of the target object/ phenomenon. It is (partly) *because* the model gets right the main causes of the phenomenon in question that it is quite empirically adequate. But since it makes several wrong assumptions, it puts the right causes in a wrong/ false setting.

Take the ideal gas model in physics. It is makes many false assumptions, but is empirically adequate with respect to the gases (at certain conditions) it is meant to give us understanding of: light gases with weak intermolecular forces at high temperature and low pressure. It is a model that takes molecules in a gas to be perfectly elastic, dimensionless, and exhibiting no intermolecular forces. Of course, no such gas exists in the real world. The model is a fiction. The model is in fact not intended to be an accurate representation of a gas. It is intended to exemplify certain features of a real gas, while masking certain other, irrelevant features, hence has several falsehoods like perfectly spherical molecules. Catherine Elgin (2009) calls these “felicitous falsehoods”. According to Elgin, such an idealized model is a strictly inaccurate representation that exemplifies – that is, it highlights, exhibits, or displays – characteristics it shares with the phenomena it purports to represent. At the same time it also *omits* irrelevant features. In doing these things, she argues, the model affords epistemic access to those features in a way a more accurate model would not. A more accurate model – in the current example, one that takes into account say, the dimensions of molecules and intermolecular forces – would be uselessly detail-rich and complex and hence reduce ease of use in the relevant domain.

How exactly do such models give us understanding? A good answer is that they do so by helping us predict and manipulate (parts of) the world. This is in fact de Regt’s (2014) account of understanding. He associates understanding with intelligibility of a theory. Intelligibility in turn is pragmatic and contextual – in the sense that the model has to get just enough things right in order to achieve the task at hand – and consists in knowing how to use the theory for prediction and manipulation. So understanding for de Regt is a *skill*. De Regt criticizes and rejects what he calls the ‘realist thesis regarding understanding’: the view that “science can provide understanding of the world only if its theories are (at least approximately) true descriptions of reality, in its observable as well as unobservable aspects.” (2014, 3) He adds, “Whether or not theories or models can be used for understanding phenomena does not depend on whether they are accurate representations of a reality underlying the phenomena,” (2014, 16). In this regard, as above, the ideal gas model helps with predictions quite successfully in certain conditions (high temperature, low pressure) – and using the model for prediction and manipulation when those conditions hold, gives us understanding.

Although the ideal gas model does not give us an accurate description and representation of reality, it can be argued that it works – helps us predict and manipulate – in a certain domain *because it gets the relevant underlying causal structure right*[6]. This is the line that Michael Strevens (2011) takes. He looks at one application of the ideal gas model: to predict the behavior of a gas in accordance to Boyle's law: the inverse proportionality of pressure to volume. He calls this 'Boylean behavior'. (Parallel discussions can be run regarding other applications of the ideal gas model.) The idea is that the idealizations and falsities in the ideal gas model do not have anything to do with the causal entailment of Boylean behavior. One can omit molecular volume, intermolecular forces, and non-elasticity of molecular interactions since these do not come in the way of the model's causal entailment of Boylean behavior. The causes relevant to Boylean behavior are the macro quantities of pressure and volume, which the model captures adequately well.

I now look at models and theories that are/ can be empirically *inadequate* and give us understanding. Here is a catalog of different kinds of models/ theories that are empirically inadequate and give us understanding in different ways:

---

[6] See John Pemberton (2005) for an argument in favor of causal idealized models in economics.

1. <u>Understanding from local explanations</u>: There are models that make false assumptions, don't get the causes right, and are also not empirically adequate in any ordinary sense. But these give us understanding by providing a good (but local and limited) explanation. E.g.: the Boltzmann atomic dumbbell model, and Prescott's representative agent models in economics.
   (Isolating models – models that are empirically inadequate but get a single cause (partly) right can also fall in this category for they give us understanding by offering a local explanation despite getting *most* of the causes wrong. But it could also be argued that they fall in the previous category of models getting causes right. Isolating models get us understanding owing to getting a single cause right, and equally owing to getting many other causes wrong – so I think they could justifiably fall in either of these two categories.)
2. <u>Understanding from proximity to the true theoretical story</u>: There are models that are predictively patently false and hence highly empirically inadequate. These give understanding owing to some proximity to the true/ accepted descriptive story about the relevant theoretical features of the target. This is premised on the views that a) a model that is more descriptively true of the theoretical features, and more empirically adequate (and presumably has other features like being visualizable, plausible etc.) gives us a great degree of understanding owing to nicely accounting for many empirical facts; and that b) understanding comes in degrees. While a model that gets more right the description of theoretical features and is highly empirically adequate can give us great understanding of the target

phenomenon, one that is empirically inadequate but somewhere in the vicinity of the true theoretical descriptive picture can give us *some* (theoretical-descriptive) understanding of the target: it is better than models that are nowhere close to the theoretical true story. But why settle for *some* understanding when we can have a fuller understanding with the help of a truer model? Because there can be situations where we are only interested in *some* understanding – when explaining things to children for instance, where the correct/ true story can be too complex. E.g.: Rutherford's model of the atom.

3. <u>Counterfactual understanding</u>: There are models that are false in several ways, not empirically adequate either, but give *counterfactual* understanding/ understanding based on possibilities. As Peter Lipton (2009) argues, understanding can be *modal*. We come to understand the world not only by saying true things about it but also by seeing how things could be and could not be. E.g.: Rutherford's model of the atom, the MIT bag model of quark confinement, and the Schelling model of racial segregation.
4. <u>Understanding from unification:</u> Theories that are false and not very empirically adequate but that do a great job of *unifying* sub-theories and (therefore) phenomena: Unification can be a very useful way to provide understanding, for instance by helping us see groups and patterns in the world and thereby reducing the overall number of phenomena.

I now look at these different understandings in turn.

Take Boltzmann's dumbbell model of a diatomic molecule. As de Regt (2009) discusses, this was in response to a difficulty that Maxwell had encountered with his formula for specific heat ratio: Maxwell found that there were anomalous gases for which the formula did not work and was very perplexed by this. The specific heat ratio depends on the number of degrees of freedom of the molecules of the gas, which in turn depends on the rotational, translational, and vibrational energies of the molecules. One of Maxwell's problems was that if one took into account all of the vibrational, translational, and rotational degrees of freedom, then one ended up with six or more degrees of freedom and the formula simply didn't work for many gases.

To solve this problem Boltzmann proposed that the anomalous gases were diatomic, with the atoms joined together by a rigid bar like in a dumbbell. This meant that the atoms couldn't vibrate (since they are rigidly held), and this reduced the number of degrees of freedom to five (three translational – along all three of the x, y, z axes, and two rotational – about the two axes perpendicular to the 'bar' joining the atoms). Of course, interatomic bonds can be stretched and broken and are not rigid. And of course Boltzmann never thought that there really were little bars joining the atoms. The molecular structure though *is* causally relevant since we're talking of degrees of freedom at the atomic and molecular level. So the model is descriptively false of the theoretical details of molecular structure, and does not get the causal structure right. It is also not empirically adequate since real gas molecules do have vibrational degrees of freedom as is shown by the fact that they emit spectral lines. By reasonable standards, not accounting for spectral lines is a significant flaw: experiment patently said diatomic gases molecules have vibration, but the model assumes they don't. I think this would, by normal standards, mean that the model is empirically inadequate; it makes wrong predictions about a significant feature in its domain. But it does get other empirical things right: it can account for the behavior of diatomic gases w.r.t specific heat ratio. But the one thing it gets wrong, it gets very wrong. Yet, it could explain Maxwell's anomalies and in this way gives us (explanatory) understanding. I will say more about this in a bit.

Another nice example of a model that is false in several ways and neither gets the causes right nor is empirically adequate, comes from Rodolfo Manuelli (1986) commenting on models of another Chicago School Nobel prize winner, Edward Prescott:

> .... consider the models Prescott surveys ... Most of them are *representative agent* models. Formally, the models assume a large number of consumers, but they are specialised by assuming also that the consumers are identical. One of the consequences of this specialisation is a very sharp prediction about the volume of trade: it is zero. If explaining observations on the volume of trade is considered essential to an analysis, this prediction is enough to dismiss such models. But if accounting for individual fluctuations beyond the component explained by aggregate fluctuations is not considered essential to understand the effects of business cycles, the abstraction is not unreasonable. A case can even be made that if what matters, in terms of utility, is the behaviour of aggregate consumption and leisure, then any model that helps explain movements in the two variables is useful in evaluating alternative policies. This usefulness is independent of the ability of the model to explain other observations. (1986, 5)

Such a representative agent model doesn't get the causes right: it omits individual consumer behavior, and there is no representative consumer as a cause. It misrepresents the causes; it substitutes a cause that in interaction with other things in the model (that are not represented 'truly' either) according to the equations of the theory will produce very roughly the same results in the targeted areas as the real causes. It is also empirically inadequate as it predicts that the volume of trade is zero, which by any reasonable standards, is very wrong.

Both of the above two models – Boltzmann's and Prescott's – make false assumptions, get a core prediction really wrong and are hence not empirically adequate in any ordinary sense of the term, and also doesn't get the relevant causes right. How do these give us understanding? They give us understanding by giving a good explanation, in some very narrow, local contexts and domains – just the contexts and domains where we *need* them to work. The Boltzmann model explains specific heat ratio for diatomic gases, and the representative agent model explains aggregate behavior.

But if they are not empirically adequate (and hence not true), isn't it just that we understand the *models* themselves but not the world? I contend that what these models help us learn about (a part of) the world is that when looked at in a certain way and with respect to some very specific aspects – like specific heat ratio or aggregate behavior of a group with respect to say, consumption of a good – it behaves *as if* the state of affairs were as described in the model. A molecule is not a dumbbell but to us, it as far specific heat ratio goes, it acts *as if* it were one, and agents are not identical, but to us, as far as aggregate consumption goes, a group behaves *as if* agents were identical. So in this sense arguably these models actually tell us things about our world.

Next, let's look at the Rutherford model of the atom. The model is entirely based on classical mechanics, and is descriptively not true and is of course highly empirically inadequate. The model doesn't get right all the theoretical features it describes about the target system – in fact it gets some core details grossly wrong. The most striking shortcoming in this regard is that it says that electrons orbit the nucleus in circles. This has the implication that electrons will lose energy continuously owing to the circular motion and spiral into the nucleus. This means atoms would be highly unstable and matter wouldn't exist the way we know it. The model is hence highly empirically inadequate. But as a result, can we say that it gives us *no* understanding of the atomic structure? That doesn't seem right. If we take descriptive truth or correctness of theoretical features as a standard for evaluating understanding, and if we take our current model as the most correct one we have, then the Rutherford model was arguably *on the path* to this most correct model in its description. After all, it's 'more' correct than it's predecessor, the plum pudding model which takes the atom to be a 'pudding' of positive charge with electrons embedded in it. The Rutherford model forms a part of a chain of continually improving models comprising the Bohr model, the Bohr-Sommerfeld model, and the modern cloud model (according to which an electron does not have a fixed position but exists as a probability wave function – or a 'cloud' – around the nucleus). How does the Rutherford model give us understanding despite its descriptive and empirical inadequacies? As Catherine Elgin (2009) says, understanding comes in degrees, and we can think of the model as a reference or starting point, and of later, more empirically adequate accounts in terms of their distance from or proximity to it. After all, the Rutherford model tells us that the positive charges in an atom are concentrated in a central nucleus containing protons and neutrons, and electrons surround it – this is a feature it shares with even the most modern model of the atom.

Elgin (2009) illustrates the important point in this context that understanding admits degrees. She considers a second grader's understanding of human evolution and notes that a belief central to her understanding might be that humans descended from apes. As her understanding becomes more sophisticated, she might hold that humans and other great apes descended from a common hominid ancestor. According to Elgin, the child's initial opinion demonstrates some understanding of evolution: it is certainly better, she says, than believing that humans descended from butterflies. So although strictly false, the child's original view does show some understanding of evolution. Her larger point is that "The growth of understanding often involves a trajectory from beliefs that, although strictly false, are in the right general neighborhood to beliefs that are closer to the truth." (2009, 325) Elgin compares the above to the typical course of development of science: "the pattern displayed by the student as he moves from the naïve view of human evolution up to the view held by the professor of evolutionary biology is the same pattern as science displays in the sequence of theories it develops." (2009, 325) Theories keep getting more and more refined and in fact theories successful at one time have routinely been shown to be strictly speaking, false by subsequent science. Each of these steps is for Elgin, a "cognitive advance". (I assume this is from the point of view of the currently accepted model or theory as the most correct one.) According to her, it would be the case that from the Rutherford model through the cloud model for instance, our understanding of the atom has got better and better over time even if, she says, we might never be able to claim that we have arrived at *the truth* about atomic structure.

Why do physics and chemistry textbooks still talk about the Rutherford model if it's so wrong? Is it to simply give students a glimpse into history? I don't think so. If the quantum mechanical model of the atom representing the electron as a wave function is the ultimate goal, then the Rutherford model nicely leads a student – through a chain of progressively improving models – to the modern quantum mechanical model. Chang's pluralism of values and goals discussed in Chapter 1 lends itself well here: we can entertain both the Rutherford model and the quantum mechanical model simultaneously, each for a different purpose: the former for understanding (in the way described above) and pedagogical reasons, and the latter for an accurate and empirically adequate model of the atom.

But the Rutherford model also gets us understanding in a different, and in my view more interesting, sense. Thanks to its wildly wrong implication of the instability of the atom, we came to see that electrons *couldn't* simply orbit the nucleus like planets around the sun. We learn that *if* the atom were like the solar system, then it wouldn't be stable. Arguably, understanding doesn't just consist in seeing what something is like, it also consists in seeing what something is *not* like, or *cannot* be like. *The Rutherford model illustrates a physical impossibility.* It also tells us that *if* the electron were such that it *did not* lose energy continuously when orbiting the nucleus, then it *could have* been the case that the atom was like the solar system. This is along the lines of Peter Lipton's

(2009) view of understanding being modal: there is understanding to be had not only from actuality, but from (im)possibility too. Oblique information – information about how things could or could not be – also gives us understanding of the world.

Consider another physics example, the MIT Bag model: it describes hadrons – particles like protons and neutrons – as ‘bags’ in which (two or three) quarks are spatially confined, forced by external pressure (similar to nucleons in the nuclear shell model). This takes into consideration the fact that quarks have never been found in isolation so far and are hence thought to be spatially confined. With the help of boundary conditions and suitable approximations, the single model parameter (bag pressure) can be adjusted to fit hadronic observables (e.g. mass and charge). Hartmann (1999, 336) observes that the predictions of the model only very modestly agree with empirical data (the pion mass according to the model turns out to be 692 MeV as opposed to the empirical value of 138 MeV). By normal empirical standards, the model fares quite badly. Are quarks really confined the way it is described in the model? We don’t know – and if empirical adequacy is a guide to truth, then very probably not. Hartmann asks why physicists entertain the model, despite its empirical shortcomings. His answer is that this and similar models provide “plausible stories” by which they enhance our understanding. The bag model is a narrative told around the formalism of the theory, and is consistent with the theory, and importantly, gives a plausible, intuitive and visualizable picture of a hadron as quarks confined in a bag. To the question, ‘How *could* quarks be spatially confined?’, this model answers, ‘Like they were in a bag’. The answer is a good one because it is easily visualizable and because it illustrates a possibility about quark confinement.

A common response to the claim that such models give us understanding is that the model is itself understandable, but it does not provide understanding of the target if it is not reasonably empirically adequate – i.e. we understand the model, but not the target phenomenon. Of course, the model is understandable since it tells us a plausible story, is intuitive, and visualizable. But it also does give us understanding of quarks' spatial confinement. As Hartmann says, "a qualitative story, which establishes an explanatory link between the fundamental theory and a model, plays an important role in model acceptance" (1999, 15) He goes on to say that all we can say is, "Something like this will probably happen inside a hadron." (1999, 10) According to him, the model gives a plausible story that plausibly relates to known mechanisms of quantum chromodynamics, the theory that is fundamental of this domain – mechanisms of non-perturbative gluonic interactions and color interaction of quarks[7]. Although empirically inadequate, the model is consistent with many other things the theory says. And anyway, given the high theory-ladenness of claims a theory such as quantum chromodynamics makes, it's not like we can somehow independently find out how quarks were *really* confined, and then compare it with what the theory says. Here again, I think modal understanding is at work: the bag model tells us one way in which hadrons *could* be confined.

To deal with the empirical shortcomings, we can follow Chang's suggestion and adopt a pluralist view and separate out the aims of understanding and empirical success: while the bag model helps with the former – particularly with respect to spatial

---

[7] See Hartmann (1999) p. 10 for a detailed discussion of this.

confinement of quarks – it fails at the latter. And that does not mean it is doomed. It just means we use to achieve an aim it helps with, not for others.

Another good example is Thomas Schelling's (1978) famous checkerboard model in economics that gives a story about racial segregation in neighborhoods. It is easy to see how neighborhoods would be racially segregated if individuals of different racial groups don't want to live around each other and have strong discriminatory preferences. But this simple scenario does not make us understand how people could be segregated even if they want or prefer mixed neighborhoods. To understand how this happens, Schelling devised the following game in which a checkerboard is taken to represent a city, with coins or other objects representing agents:

1. The checkerboard is filled with say, nickels and dimes representing two different races.
2. Squares adjacent to a coin form its neighborhood. So interior (non-edge) coins can have eight neighbors, non-corner edge coins can have five, and corner ones, three.
3. Rules are devised to determine if a coin – representing an agent – is 'happy' with its neighbors. If unhappy, it moves to a different, vacant location on the board.
4. The game continues until 'agents' have their preferences met according to the happiness rule in place, and this state represents the final, equilibrium outcome.

The outcome primarily depends on the happiness rule. If the rule is such that the threshold tolerance for the other race is so low that 'agents' want *all* their immediate neighbors to be of their own race, then not surprisingly this leads to an outcome of complete segregation. The interesting part is playing around with other rules. Schelling found that even rules expressing small preferences – like "I want at least a third of my neighbors to be of the same race" – could result in complete segregation.

What's worth noting about this model is that it does not represent any real city. Moreover, it makes *no* empirical predictions since it doesn't deal with any real situation. It deals with a *possible* situation. What it does is far more subtle and interesting. As Aydinonat Emrah (2007) notes, the model is constructed to see how certain individual mechanisms (i.e. individual tendencies to avoid a minority status) *may* interact under certain conditions. As he says,

> Maybe the only aspect of the model that is familiar to us (i.e. represented in the model) is these individual mechanisms: we know that individuals tend to avoid a minority status and need to belong to a certain group. In fact, it is due to the familiarity of the individual mechanisms presented in the model that we tend to think what happens in the checkerboard city may happen in the real world. (2007, 439)

He goes on to point out that the model makes a *plausible conjecture* based on our familiarity with people and their preferences, that segregation may be the unintended consequence of even mild discriminatory preferences. But the checkerboard model may be plausible and interesting, yet it does not tell us whether mild discriminatory preferences bring about segregation in the real world. But it quite amazingly opens our mind to a previously unimagined possibility – who'd have thought that such mild racial preferences can possibly lead to complete segregation? Here then is a case of a celebrated model in economics – celebrated for providing insight into and understanding of racial segregation – where empirical adequacy simply does not figure.

There are several advantages of models that are empirically false or even empirically sterile, not making any predictions. First, as Hartmann pointed out, they give us plausible stories to think of the relevant phenomena in terms of. Second, as in the case of the Rutherford model, we can think of more empirically adequate accounts in terms of their distance from or proximity to an empirically inadequate model which can serve as a reference or starting point. Third, these models give us counterfactual 'what-if' scenarios to think about and understand the world (Lipton, Emrah). As would apply to many of the examples I've discussed, empirically inadequate models (and theories or explanations) that are adequate in other ways (giving plausible stories for instance) can be thought of as generating understanding via providing us with *possible* scenarios rather than actual ones. The checkerboard model for instance lends insight into how individual tendencies – that we are familiar with and know to exist – may work together. Lipton's words seem to ring true here: "Information about a possible mechanism may give oblique information about the actual mechanism (telling us something about a class into which it falls) by describing a mechanism in the same class, namely the class of mechanism that can generate this phenomenon." (2009, 51) Surely, situating a real world case within a grid of associated possibilities increases our understanding of that case by illuminating the distance/proximity of the case under study from its cousins.

Lastly, I look at understanding and explanation. It is commonly held that explanations in science help us understand why something is the way it is. Peter Lipton (2001) for instance argues that understanding is the intrinsic (as opposed to instrumental) good that comes out of explanation. The question I'm concerned with is: in order to gain understanding, does the explanans of the explanation we grasp have to be true? The answer for many is unsurprisingly, a resounding yes. But there are explanations in science whose explanantia are far from true or even empirically adequate. And arguably, they do give us understanding. I will consider one such type of explanations – unifying explanations. Explanations that unify a number of apparently diverse phenomena into fewer groups increase our understanding of the world.

If unification is a source of understanding, just what more exactly is it doing for us that warrants that label? There are a number of answers on offer. Empirical adequacy is not necessary for any.

1. A unification gives us a picture of the world 'as a whole': the understanding it provides is global as opposed to local. Unifying explanations may not increase our understanding of independent phenomena, but they do increase our understanding of phenomena overall, and the world at large, by bringing together apparently diverse phenomena under one or a few theories. (Friedman, 1980)

2. Unification reduces the number of independent phenomena and laws in our theories. For Friedman, for instance, this reduction is the very "essence of scientific explanation".
3. Unifying explanations can show widespread patterns in nature. Getting the right unifications can carve the domain of empirical laws in to the correct classes in which they belong[8] – perhaps not 'correct' in some grand, metaphysical sense, but in an ordinary sense, applying to the observable world.
4. As Mary Morgan (2010) points out in her work on the travel of facts and techniques from one domain to another, it can be extremely useful to see that the laws all grouped together under the same unifying claim are similar in significant ways. It allows us to use similar methods of study, modeling strategies, approximation techniques and the like and it suggests analogous predictions to look for from one domain to another.

Note that a unifying theory can succeed at all of these jobs without being as empirically adequate as the theories that are unified. First, if the job is to see the whole all at once, or to reduce the number of independent pieces, it may not matter that the unifying theory makes correct empirical predictions. Second, we can see and appreciate the pattern even if each individual piece is not entirely accurately represented in the unification and departs in various ways from it. And seeing things together that are very much alike is a way of understanding them that is in no way dependent on the truth of the

[8] This is similar to Pierre Duhem's (1906/1954) idea of natural classification (which I discuss in depth in Chapter 6) although Duhem thought theories classify laws in a way that progressively reflects an underlying, metaphysical, natural classification.

vehicle that unites them. Lastly, when it comes to borrowing techniques, looking for predictions in one domain analogous to those already established in another, and the like, it is the analogies among the unified sub-theories that matter, not the empirical adequacy of the unifying theory. The unification may be substantially less empirically adequate than the ones being unified. That's because as above, the unifying theory may not get any of the unified subtheories *exactly* right – but *right enough* to help us see a pattern – and so may make many empirical mistakes in the domain of each. So empirical adequacy is a poor guide to understanding via unification. *It doesn't matter if the unifier is as empirically adequate as the things being unified.*

Again, one might say that if a unifying theory is false, while the theory itself might be understandable – owing to being 'neat', organized, and structured, it does not give us understanding of the phenomena unified or the world at large. But again, this seems unreasonable. Even if the world is messy and far from unified at the level of 'essences' or 'noumena', unification arguably gives us understanding of (at least) the phenomenal world – our world, the world we live in and experience. Also once again, I contend that unification gives us an 'as-if' understanding: even if the world were 'really' messy and not unified, in certain contexts in which *we* work and interact with the world – for instance if methods/ techniques used in one domain also work in another – the world is as if phenomena were unified.

2.3 Conclusion

Van Fraassen (1980) claims that empirical adequacy is *the* semantic virtue that theories (and models) should have with respect to how they relate to the world. But I contend that empirical adequacy does not capture the multitude of ways in which a theory or model is expected to relate to the world in scientific practice. If all one is interested in is truth about observable phenomena (and about unobservable phenomena as well for the realists), then sure, one can fervently pursue empirical adequacy. And doing this can also certainly bring about understanding. But stating true facts is not the only way that theories and models can give us understanding of and insight into the workings of the world – and to this end they can reasonably be expected to be adequate in different degrees and with respect to different empirical and non-empirical factors. Brown's pragmatic functionalism (as discussed in Chapter 1) applies well here: depending on our pragmatic goals – here understanding – we decide what the optimal balance of values (intelligibility, plausibility, story-telling prowess, explanatory power, local and limited predictive power etc.) and empirical adequacy is. Singularly, empirical adequacy – however charitably defined – is an inadequate standard of evaluating all of scientific practice. Scientists have different aims and work on different problems in different contexts and settings and often forego empirical adequacy in favor of other values like plausibility, possibility, visualizability and explanatory and unifying power.

This set of claims amounts to a form of quite radical pluralism, without becoming full-fledged relativism. The focus is primarily on the aim or task at hand. And whatever theory or model helps achieve a specific aim(s), it must be fully accepted and entertained for that specific purpose. So it's not the case that anything goes; only those stories and models go that help achieve the task(s) at hand. And the more specified the task (gaining a particular kind of understanding – predictive, visualizable etc.; answering a specific 'why' or 'how' question, and so on), the clearer the purpose of the theory or model used, and lesser the room for the allegation of relativism. But of course, I am not suggesting that it is possible to carry on scientific practice without any regard to empirical factors. Experimentation has been the backbone of scientific practice as a whole and will continue to be so. But there is evidence that values other than empirical adequacy have also been backbones of scientific practice. And there have always been, and will continue to be, pockets of scientific practice that involve very little or no empirical considerations – and this doesn't make them any less scientific.

Portions of Chapters 1 and 2 are being prepared for publication by Nancy Cartwright and myself and to appear as

Bhakthavatsalam, Sindhuja; Cartwright, Nancy. 'What's so good about empirical adequacy?'

References

- Brown, Matthew J. (2013) 'Values in Science Beyond Underdetermination and Inductive Risk' Philosophy of Science 80 829– 839

- Chang, Hasok (2001) 'How to take Realism beyond Foot-Stamping' *Philosophy* Vol 76 No. 295 Cambridge University Press 5–30.

- De Regt, Henk (2009) 'Understanding and Scientific Explanation' in de Regt, Henk; Leonelli, Sabina; and Eigner, Kai eds., *Scientific Understanding: Philosophical Perspectives* University of Pittsburgh Press 21–42

- De Regt, Henk; Leonelli, Sabina; and Eigner, Kai (2009) 'Focusing on Scientific Understanding' in de Regt, Henk; Leonelli, Sabina; and Eigner, Kai eds., *Scientific Understanding: Philosophical Perspectives* University of Pittsburgh Press 64–82

- De Regt, Henk (2014) 'Scientific Understanding: Truth or Dare' *Synthese*

- Duhem, Pierre (1906) *The Aim and Structure of Physical Theory* (Repr. 1954) (Philip P. Wiener, Trans.) Princeton University Press

- Elgin, Catherine Z. (2009) 'Is Understanding Factive?' in Duncan Pritchard, Allan Miller, Adrian Hadock, eds. *Epistemic Value* Oxford University Press 332-330

- Emrah, Aydinonat (2007) 'Models, conjectures and exploration: an analysis of Schelling's checkerboard model of residential segregation' *Journal of Economic Methodology* 429-454

- Friedman, Michael (1974) 'Explanation and Scientific Understanding' *The Journal of Philosophy* Vol. 71 No. 1 5-19

- Hartmann, Stephan (1999) 'Models and stories in hadron physics' 326-346 in Morrison, Margeret and Morgan, Mary S. *eds., Models as Mediators* Cambridge University Press 326-346
- Hempel, Carl G. (1965) *Aspects of scientific explanation and other essays in the philosophy of science*. New York: Free Press

- Lipton, Peter (2001) 'What good is an Explanation?' in G. Hon & S. Rakover (eds.), *Explanation*. Springer Netherlands 43-59

- Lipton, Peter (2009) 'Understanding without Explanation' in de Regt, Henk; Leonelli, Sabina; and Eigner, Kai eds., *Scientific Understanding: Philosophical Perspectives* University of Pittsburgh Press 43-63

- Manuelli, Rodolfo (1986) Quarterly Review *Federal Reserve Bank of Minneapolis* Vol. 10 No. 4

- Morgan, Mary (2010) 'Traveling Facts' in *How Well Do Facts Travel?*

The Dissemination of Reliable Knowledge in Howlett, Peter; Morgan, Mary S. eds., Cambridge University Press

- Pemberton, John (2005) 'Why ideals in economics have limited use' *Poznan Studies in the Philosophy of the Sciences and the Humanities*

- Schelling, Thomas. C. (1978) 'Micromotives and Macrobehavior' London and New York: W.W. Norton

- Strevens, Michael (2008) *Depth: An Account of Scientific Explanation* Harvard University Press

- Trout, J.D. (2007) 'The psychology of scientific explanation' *Philosophy Compass* 2 564-91

- Van Frassen, Bas (1980) *The Scientific Image* Oxford University Press

3. Ontological Plausibility and Scientific Understanding

3.1 Overview

In his 2001 paper 'How to take realism beyond foot stamping' Hasok Chang argues that an empiricist realism[9] - a realism that considers empirical success of scientific theories to be epistemically valuable and attempts to establish their truth based on that success – is not viable by itself, and attempts to arrive at a more reasonable version of scientific realism, going beyond empiricism. As we will see though, the focus is not on truth: his is a realism that is not truth-centric. Chang proposes an entirely non-empirical criterion for evaluating our systems of knowledge. He calls this the criterion of "ontological plausibility". For Chang, the motivation for this comes from our frequent inability to abandon certain intuitive (non-empirical) beliefs we have about reality – for their denial would be ontologically implausible to us. These unshakable beliefs, he calls "ontological principles": rationally compelling guiding principles for our comprehension of reality. He advocates adopting such principles as serious criteria for evaluating our knowledge claims about the world, over and above empirical considerations. An example would be what he calls the "principle of single value": a measurable physical property

[9] Although this label might sound like a contradiction in terms – since empiricism is well known for its attack on realism – Chang gives this name to empiricist positions that have at the same time been realist: positions that take empirical success of theories to indicate, as he quotes John Worrall, that their "descriptions of an underlying reality are accurate, or at any rate 'essentially' or 'approximately' accurate". Chang cites one such position, from Grover Maxwell's paper, "The Ontological Status of Theoretical entities" where Maxwell says, "Since a well confirmed theory (plus, perhaps other well confirmed sentences) entails that there are lines of force, lines of force exist". It is claims such as this that Chang has in mind when he uses the term 'empiricist realism'.

can have one and only one value in the same situation at any given time and place[10]. Denying this, says Chang, seems to be *unintelligible* and result in incomprehension. Therefore any encounter of the denial of this principle should tell us that something is amiss, for to us, the denial is unacceptable. A realism that epistemically values and pursues these principles – and takes them as some kind of standard for evaluating validity of scientific claims – he calls *plausibility realism*.

My goals here are to 1. Situate Chang in the values-in-science domain and look at plausibility and intelligibility as values and how they fare against empirical adequacy, 2. Evaluate Chang's account of scientific understanding. I conclude that the 'plausibility account of understanding' is not sustainable without having truth in the picture. As in Chapter 1, following distinctions from Larry Laudan (2004), Heather Douglas (2009) argues that epistemically, while empirical adequacy is a "minimal criterion", there are other "ideal desiderata" or values like simplicity, scope etc. I bring to attention that Chang does not make a distinction between plausibility and intelligibility. I argue for a difference between the two in terms of minimal criteria – desirable values distinction as above: I contend that while intelligibility is a minimum epistemic criterion, plausibility is to be seen as a value. But I argue that plausibility alone – without any regard for empirical adequacy – is not really a desirable value, for contrary to Chang I don't think it really helps us with understanding reality.

[10] Chang is quick to offer a clarification with respect to quantum mechanics, though only in the later paper: he says, "Even quantum mechanics, even in its Copenhagen Interpretation, stops short of saying that a particle can have multiple positions at once; a particle can only have non-zero probabilities of detection in various places."

3.2 Ontological plausibility and Scientific Understanding

3.2.1 Background and Exegesis

In *How to take realism beyond foot stamping*, Chang starts by dismissing empiricist realism on the grounds of pessimistic induction: we have a “graveyard” of previously empirically successful theories, which have eventually been shown to fail. In response to this, Chang points out, realists have tried to separate parts of a theory that are genuinely responsible for its success from other parts that may be discarded once it is realized that they were inessential to begin with. While it is not clear in the first place if this kind of separation is feasible; the more important point for Chang is that the core argument for any empiricist realism will have to involve defining truth in terms of empirical success according to which empirical success *means* truth. Realism would indeed be true, but would be tautologically true – it would boil down to something as trivial as “realism is empiricist realism” – and would essentially add nothing to basic empiricism.

Chang proposes that we move beyond empiricism and take for our guidance in theory choice and evaluating our knowledge of the world, the criterion of *ontological plausibility*. This criterion consists in non-empirical, plausibility-based principles he calls 'ontological principles' which are “regarded as essential features of reality in the relevant epistemic community” (11). These principles are such that denying them would seem unintelligible to us: “Ontological principles are the basis of intelligibility in any account of reality; the denial of an ontological principle strikes one as more nonsensical than false**.**” (11) They thus seem cognitively *necessary* to us to make sense of, and *understand* the world. Chang's “alternate form of realism” which pursues these ontological principles is claimed to avoid the above problem of empiricist realism by turning to principles that are altogether non-empirical.

Chang provides instances from the history of science that evidence the significance given to such plausibility considerations by widely regarded scientists, thereby giving credence to his views: Einstein was unable to accept indeterminism in quantum mechanics on plausibility grounds; Leibniz proposed a law of continuity again on plausibility grounds, and so on. To further demonstrate the important role played by such principles in doing science, he narrates the story of how Henri Victor Regnault, in trying to arrive at the most appropriate substance for using in a thermometer employed (tacitly) the principle of single value: while mercury and alcohol did not result in consistent results for the measured value of temperature (in the same circumstances), the air thermometer did- which led him to conclude that the air thermometer was the most suitable. Chang’s point is that Regnault had to assume that if temperature were a real property then it *had* to have a single value. Denying this would only result in utter incomprehension. If one were to report, for instance, that the temperature of an object is both 5 degrees and 10 degrees (in some chosen units) at the same place and time, that would be unintelligible (and implausible – Chang doesn't make a distinction here). Chang further says, such a statement does not even qualify to be assigned the truth value “false”: it simply does not make any sense.

What exactly is the source and nature of these ontological principles? And how reliable are they? Chang likens these principles to Kant's synthetic a priori – they are *intuitive* and non-empirical. It’s not the case as he explains, that we go around empirically testing these principles. For Chang, what this illustrates is that there are some deeply ingrained beliefs in us that go over and above empirical considerations, and yet seem necessary in helping us make sense of reality. Of course we might be proved wrong about these as Chang admits, but he makes the point that some principles being false by itself does not discredit the motivation to rely on plausibility as an epistemic virtue – just the way that failure of some empirical methods is generally not taken to discredit empiricism itself. In addition to the principle of single value, examples of ontological principles he offers are the “principle of physicalism”: all real objects are located in space and time; the principle of 'no miracles': “if there are regularities in nature, they cannot be suspended on isolated occasions to allow inexplicable happenings ('miracles')”; the principle of “no infinities”: a real physical quantity cannot take on infinite values, and conservation: the belief that there is something constant in the universe that cannot be created or destroyed. (Chang admits however, that he is less certain of the latter two being ontological principles than he is about the former two.)

An important point of departure from Kantian synthetic a priori though is that Chang allows ontological principles to be revisable – they are not absolutely universal and unchanging. (Here he gives the example of Euclidean geometry: what was once thought to be universal turned out not to be so.) But although ontological principles are revisable according to Chang, it is not the case that we have *no* principles that are stable and reliable: he believes that we do have some relatively secure principles, “principles to which most people will give firm assent after careful consideration” (16). Examples include the principle of single value or the principle of physicalism. According to Chang, we should start with such principles. These are to be distinguished from mere “ontological conjectures or prejudices”. Chang gives two clear characteristics of legitimate ontological principles:

1. Their denial seems utterly unintelligible (so something like determinism wouldn't qualify: for its denial (indeterminism) is not unintelligible to us in the way the denial of the principle of single value is)
2. They cannot be empirically tested, whatsoever. Talking of the principle of single value: “If someone would try to support this principle by going around with a measuring instrument and showing that she always obtains a single value of a certain quantity at a given time, we would regard it as a waste of time.” (12)

Ontological principles – principles that meet these two criteria – are not mere opinions and are much more fundamental than what is given to us in ordinary material experience, which is why to test them empirically does not make sense. Although history tells us there have been a number of episodes of adherence to false ontologies – Newton's opponents who proposed their “ontological principle” of contact action, for instance, Chang defends his ontological principles against accusations of being more uncertain, subjective and idiosyncratic than empirical claims. Chang argues that similar problems of fallibility exist with empirical beliefs as well: our senses are known to fail us often; different observers often disagree while observing the same (external) thing; and so on. He believes that just as we allow ourselves to improve and standardize empirical observations (without easily abandoning them if they are unsatisfactory), we can indeed also do the same with ontological principles with "sufficient effort and creativity" (13) without entirely giving up on them, suspecting to be ultimately led to a "relativistic morass" (14): "If fallibilist empiricism is allowed to roam free, there is no justice in outlawing ontology because it confesses to be fallible." (15) Further, he insists that there is no good reason for ontological plausibility to be subordinate to empiricism:

> **C**: If some people choose to do their best to reform their ontological beliefs to fit with the current best scientific theories, I have no particular objections to that. However, if the argument is that all of us have a duty to make such efforts to give up our ontological principles to state-of-the-art science, I do have some objections. (2001, 14)

So where does empirical adequacy stand in relation to these principles? According to Chang, empirically adequate theories that are not ontologically plausible would be treated merely instrumentally for heuristic purposes, but not for understanding or explanation. Chang notes that Newton himself had a similar take on his own equations of gravitation since action-at-a-distance was unintelligible and ontologically implausible to him. Importantly, Chang insists we treat plausibility criteria for evaluating scientific claims on at least the same footing – if not higher – as empirical criteria.

Would an empirically *in*adequate theory be accepted if it were ontologically plausible though? Chang's above quote (C) is ambiguous in this regard. By 'state-of-the-art science', Chang means empirically successful science. Now does he mean that empirical success is alone not a good enough standard to judge theories and that they *also, in addition* need to be ontologically plausible, or does he mean that it is acceptable to strive for and accept ontologically plausible theories and principles (for the purpose of understanding) even if they are not empirically successful? Just like he says that an empirically adequate theory that is not ontologically plausible may be used for say prediction but not understanding, does he also hold that an empirically inadequate theory may not serve the purpose of prediction, but may serve the purpose of understanding by being ontologically plausible? He asserts, "Sticking to ontological principles may seem like unnecessary conservatism, but always following the latest scientific opinion is unprincipled fadishness." (14) Clearly according to Chang, those theories that are ontologically more plausible than others are more valuable for the purpose of understanding the world. (There is also the question of whether ontological plausibility is in fact a good criterion for understanding, which I discuss in the next section.) He says, "…given a set of theories that are empirically as good as each other, if one is ontologically more plausible than all the others (and hence capable of giving the best explanation), then we must treat that one as the most favourable alternative."(28) Here it looks like Chang thinks that plausibility can break the tie between multiple empirically adequate alternatives. But can plausibility sometimes trump empirical adequacy? Chang doesn’t answer this directly, but the following quote at the end of the paper seems to suggest that Chang’s answer is in the affirmative:

> To summarize: If a system of knowledge is empirically successful, it will help us function; if it is ontologically plausible, it will help us understand; if it is both, that is even more wonderful. It is not clear what else we could reasonably demand or claim. To insist on some 'truth' beyond utility and intelligibility would only amount to foot-stamping. (2001, 30)

So Chang’s overall view seems to be that different theories (or other bodies of knowledge) in science can be used for different purposes: a theory that is empirically adequate but ontologically implausible can be used for say, prediction, and a theory that is ontologically plausible but empirically inadequate, can be used for the purpose of understanding. (And it’s “wonderful” if it’s both.) But as I discuss in the next section, there is a problem with this view.

What about truth in this story? Do ontological principles bear any relation to truths about ontology? Is ontological plausibility somehow a precondition for truth? Chang denies this, for we haven't established the truth of ontological principles themselves in the first place. Even if it were a precondition for truth, he says, such a truth would not serve any purpose not already served by empirically adequate theories. Thus, Chang's approach is to steer clear of truth altogether. The goal in his realism is to pursue intelligibility independent of truth or empirical adequacy. For him, ontological plausibility is a precondition for intelligibility irrespective of whether our ontological principles are true. Importantly, what ontological plausibility considerations help afford is *understanding*: in solving a scientific problem, by employing only principles that are ontologically plausible to us, we ensure that we *understand* the phenomena at hand as opposed to approaching the problem merely algorithmically, for instance. A case in point for Chang is how Newton's opponents considered the concept of action at a distance to be implausible – it didn't, in their judgment, give understanding. As above, Chang says that although the anti-Newtonians' principle of contact action may have been misguided, we can appreciate their demand for understanding. And also, since explaining is central to understanding, adopting ontological principles is desirable, for they give us better explanations: ontologically plausible explanations are better than ontologically implausible ones. Chang opines that understanding and explanation are of crucial importance in scientific practice and that one should not always merely aim at description and prediction. His commitment is to explanations and understanding based on ontological plausibility considerations and he is thus ready to "give himself up as a hostage to metaphysics". He acknowledges though, that different practitioners of science might have different priorities and may not value explanation and understanding within science.

But doesn't such a pursuit of understanding run the risk of subjectivity? What about the uncertainties of metaphysics? Doesn't it ground understanding in the personal metaphysical predispositions of scientists? As one of the responses to this worry, Chang warns against being dogmatic about ontological principles – much as Einstein was when he declared God didn't play dice – and asks that we be open to revision. Although Chang admits that is an open question whether we can have a large enough list of genuinely helpful ontological principles, he insists that it is a goal to be pursued to the best of our abilities – just as some realists insist that we pursue truth even if we can't be sure of whether and to what extent we can attain it. He adds that empirical claims are no more immune to being fallible anyway, so plausibility claims should not be held to a higher standard. And just as we keep working to improve our empirical standards and methods, we can work to improve the quality and quantity of our ontological principles.

Further, Chang wants to be *realist* about scientific theories/ claims/ principles that are based on ontological plausibility considerations. But what good is a realism that is based entirely on what is intelligible and plausible to *us*? Where is truth in this whole story? How does this story qualify to be a species of realism at all? Chang's response is that plausibility realism is realism since it goes beyond description and prediction, to understanding and explanation. But what does this achieve other than probably a subjective satisfaction? What good is a realism without truth? Here is where Chang makes the important point that for him, a Truth with a capital 'T' – by which I take him to mean some kind of an objective truth about a reality underlying sensible phenomena – is an illusory goal: it is attainable neither by our senses nor by a priori reasoning. He goes on to say,

> In fact, even if we somehow came into possession of it, we could not be sure that we had it. Insistence on this kind of Truth is only an open invitation to scepticism, as futile as the investigation of Kantian noumena. Plausibility realism may be seen as the pursuit of truth, but only in a subject-relative sense of truth (2001, 26)

"If we were to bet", he says, an ontologically plausible theory would be more likely to be true than one that is not; but he is happy to just focus on pursuing intelligibility, and in its train, explanation and understanding, without getting distracted by such "Truth-gambling". For him, as long as we are intellectually satiated with a sense of understanding and having explained something, a further goal of truth (of any kind) is rendered moot.

Chang winds up with the point that ontological principles are often employed in active empirical science itself. To illustrate this, he brings to our attention Ian Hacking's discussion of a situation where several microscopes employed to study the structure of dense body in blood platelets revealed one and the same structure. In deciding which of two possible, empirically equivalent explanations we should accept: "(i) The real microscopic structure of the specimen is as shown by the microscopes. (ii) The real structure is not as observed, but a complex set of conspiratory deceptions make it appear so" (2001, 28) Hacking insists that it would be a "preposterous coincidence" if all the microscopes, despite each showing the exact same visual structure, were not in fact revealing the actual structure of the bodies. For Chang, this is a classic case of the use of an ontological principle: it was extremely hard for Hacking to believe that "nature would behave in such a way as to deceive us intentionally, and with the same result time and again". (2001, 27) Chang calls this ontological principle the principle of "cosmic honesty" – the principle that nature wouldn't deceive us multiple times in such a situation – and goes on to say that there exist many situations like this where what is thought of as empiricist realism is actually plausibility realism.

3.2.2 Reflection and critique

Recall Douglas' distinction between minimal criteria and ideal desiderata from Chapter 1. While I think what falls in each category is debatable (and I have argued contrary to Douglas, that empirical adequacy need not be a minimal criterion), one value that I think definitely falls in the category of minimal criteria is intelligibility the way Chang construes it. I.e., according to my redefinition of minimal criteria – that they are those that always need to be met – I contend that intelligibility is a minimal criterion. But we should first see that in the context of Chang, plausibility and intelligibility come apart (and this difference will become even more important for the discussion of his next paper) – and Chang does not make this distinction. The difference, going by Chang's own construal of intelligibility, is that something is unintelligible when it is incomprehensible, nonsensical; but something is implausible when it is hard for us to imagine or believe it, but is not altogether incomprehensible or nonsensical. And plausibility can be a continuum: something can be more or less plausible than something else; whereas intelligibility – going by Chang's example of the principle of single value – is usually binary: something is intelligible or it is not.

Let me start making my case for this distinction by noting that Chang's list consists of principles that don't all seem to go together. Although in a bid to make them objective he wants to be very strict about the criteria of denial-being-unintelligible and empirically untestable and argues that the idea of determinism for instance is not an ontological principle since its denial is perfectly intelligible (2001, 15) – we can perfectly well imagine a non-deterministic universe – it is clear that only some of his enlisted principles are based on intelligibility. There is clearly a disconnect between Chang's criteria – of denial being unintelligible and being empirically untestable – and some of his own examples, for they seem to violate these criteria. From among the principles on his list, while the denial of the principle of single value and the principle of physicalism seem outright nonsensical/ absurd – hence unintelligible, the denial of the principle of 'no miracles' or the principle of cosmic honesty are clearly not unintelligible: they are perfectly imaginable and certainly not incomprehensible, and hence perfectly intelligible. But they do seem *implausible.*

In fact, that the denial of the principle of cosmic honesty is not unintelligible is clearly demonstrated by Ronald Giere's (2006) attempts to falsify it and fully support its denial. A key feature of his perspectivist philosophy is that no matter how many different ways we have of studying the same thing that corroborate each other, they are all still inescapably perspectival: i.e. what each of them reveals is relative to the relevant perspective of the theory/ instrument. In the context of Hacking's example that Chang cites in favor of the principle of cosmic honesty, perspectivism implies that every microscope's image is relative to each of the respective microscope's perspective; no image guarantees the actual, real underlying microscopic structure. While the denial of the principle is thus not simply unintelligible, one can however argue that it is certainly implausible. (Giere's arguments are of course based on considerations other than plausibility.) Chang thus needs to make a distinction between plausibility and intelligibility.

Let us pause here and ask, *what* is it that Chang is insisting on, about our attitude toward ontological principles? The answer is twofold. First, he is prescriptive: he says that we should place great cognitive value on these principles – for they lead to understanding and explanation – and employ them in making decisions about theory choice. Second, when we *do* employ these principles, we often do so without realizing their rationalist rather than empiricist grounding, and so Chang wants us to *acknowledge* their use in the practice of science so we become aware of and realize the significant role that non-empirical, rational factors play in scientific theorizing. Let's first look at the second point. Chang's point is well taken – this realization of the use of rationality-based ontological principles – be they intelligibility-based ones or plausibility-based ones – surely gives us a better understanding of scientific practice and would help us move beyond blind empiricism. A hardcore empiricist/ empiricist realist might indeed not realize that simple measurement of temperature or the study of the microscopic structure of platelets relies on an entirely non-empirical, intelligibility/ plausibility based principle, that of single value and cosmic honesty respectively, and so Chang's insistence that we acknowledge the rational basis of such principles is reasonable. This acknowledgment can be particularly revealing in the case of plausibility-based principles. For instance, although many scientists routinely use the principle of cosmic honesty, realizing that it is a plausibility-based principle and that it plays a major role in scientific decision-making could be a significant revelation to someone who naively claims that empirical evidence should be the principal word. Since denial of this principle can be perfectly rationally entertained – one can perfectly well comprehend the possibility of different microscopes producing images of an object that are consistent with each other, yet each not revealing anything about the actual structure of the object – the fact that we do infer the contrary, based on *plausibility* considerations – is indeed something worth taking note of: it will help realize that, as Chang says, “a lot of what's out there that looks like empiricist truth realism is in fact plausibility realism” (2001, 27)

Turning to the first point now – prescription of ontological principles – given the distinction I have made between intelligibility-based and plausibility-based principles, the former seem so obvious and fundamentally necessary to our rationality that they don't need to be vehemently prescribed. I reiterate that intelligibility is, in Douglas' terms, a “baseline criterion”. The single value principle and the principle of physicalism for instance, are principles that empiricists, rationalists, 'plausibilists' and everyone else will vouch for. I don’t imagine that anyone in the business of science would claim that a physical quantity like temperature can have two values at the same point in space and time, or that physical bodies don’t exist in space and time. So their employment by different scientists in different scenarios (Regnault for instance, as discussed above) doesn't come as a surprise and need not be forcefully prescribed, since they are anyway already in practice. Acknowledging that intelligibility-based principles are pervasive in scientific practice (due to which I have argued, they need not be prescribed), is their use justified? The answer I think is yes, but for now all I can say is that that is because they are so basic to and in fact constitutive of our rationality – we cannot function without them. In the next chapter I offer a pragmatic explanation and justification of this universal adoption of these intelligibility principles based on Chang’s later work on epistemic activities and pragmatic rationality.

What about the latter, plausibility-based principles, like the principle of cosmic honesty or the principle of no-miracles? As I’ve argued, these are not as rationally fundamental as the intelligibility-based ones: we can perfectly well imagine their denial. Should these be prescribed in scientific practice? Keeping with Chang’s rationalist spirit the answer would be yes. For instance, Einstein’s belief that “God doesn’t play dice” – which is clearly based on plausibility and not intelligibility considerations – conflicted with (quantum mechanical) empirical evidence (which was decidedly probabilistic) and a Changian plausibility realist would argue in favor of the rational motivation behind such beliefs (even if they don’t support the belief itself) in scientific practice. Importantly, intelligibility based principles alone seem to be insufficient. If one wants rationally intuitive considerations about ontology to play a significant role in science then one had better be more flexible on the one hand, not restricting oneself to only principles whose denial is strictly unintelligible; and take risks on the other, placing faith even on some of those principles that aren't obviously (rationally) necessary to us but whose denial is nonetheless implausible. Chang's criteria are unreasonable and overly restrictive for his own good – and as I’ve shown, he doesn’t stick to those criteria himself in his examples of ontological principles.

To elaborate, consider a situation where we have a theory, no aspect of which is unintelligible in Chang's sense, but let's say something in it seems particularly implausible. If we hold on to Chang's rigid criteria of the principle being empirically untestable and its denial being unintelligible, then we should not take this implausibility seriously. And if this becomes part of our practice, a lot of theories might slip by without going through sufficient scrutiny with regard to rational intuition concerning ontology. Relying on principles based strictly on intelligibility wouldn't be good for a Changian since they seem too few and far in between – Chang's list is quite short and his hope that "more can be found" is anything but promising. So if we don't embrace and endorse plausibility principles in addition to the intelligibility ones, we wouldn't have enough rationalist checkpoints for theories – but this is in fact counter to the spirit of Chang's philosophy. Plausibility-based principles might face the problem of being subjective and idiosyncratic, but we could find ways to deal with that – we might for instance endorse just those principles that are relatively stable across several people.

I've argued that from within Chang's perspective – *if* we want to meet our rational ontological intuitions in scientific practice (at least for the purpose of understanding) – then plausibility should be valued in addition to intelligibility. But is such a goal warranted at all? The answer, I think, depends on the context and purpose. First, if the purpose is to get to the (approximate) truth about the world, are intuitive plausibility considerations a way to that? Probably not. Prima facie, there's no reason to think that our rational intuitions do or tend to latch on to the truth about the nature of the world. And Chang himself concedes that he doesn't establish ontological principles to be true.

Chang's view is of course removed from truth considerations – his position is not committed to any kind of truth. But for this reason in fact, I think Chang should not advance a version of *realism* based on plausibility. Chang admits himself that plausibility realism is not a 'truth realism' and doesn't have anything to do with truth, but rather with understanding and explanation. But why label it as a species of realism at all then? And anyway, Chang doesn't *need* to appeal to any form of realism to make his arguments work: his central idea that ontological plausibility is an important factor for understanding stands on its own as a separate point, not having to do with the issue of realism (though I go on to argue below that such an account of understanding is not tenable without truth in the picture). If Chang's realism involved truth, it would deserve the label of realism and would have also helped his account of understanding in my view. But since his realism does not involve truth at all, I think it is neither correct nor required (for the arguments he makes) to call his view a species of realism.

And this brings us to understanding: if the purpose is understanding the world – regardless of any truth concerns – then are intuitive plausibility considerations to be valued and prescribed? The answer depends, I think, on the kind of understanding we want. If ontology being intelligible/ plausible to us is what makes us understand it, then I don't think ontological principles should be prescribed for the sake of understanding without truth in the picture, and here's why. According to the goal of achieving understanding by coming up with theories and principles that make ontology come out to be ontologically plausible/ intelligible – call this the plausibility account of understanding – what we're essentially saying is that for an ontology to be understandable, it has to have certain plausible/ intelligible features. Or to be more precise, it should *not* have certain implausible/ unintelligible features like the feature of a having a single point in space and time that has multiple values for temperature, or the feature of deceiving us about the microscopic structure of a certain object under study (in reference to the Hacking scenario discussed earlier). We might be able to come up with theories and principles that construct such an ontology, but we can't be sure that the ontology we are trying to understand – of our world – matches the ontology of such theories: and if it doesn't, then we're not understanding the ontology we want to understand – what we're probably understanding is a make-believe ontology within the theory that we have bestowed these (plausible) features upon. In this plausibility account of understanding then, it is not possible to understand without truth: it is not possible to understand ontology with the help of a theory that does not in some sense *truly* capture its features. (Truth) realism is built into this account of understanding – without realism this account of understanding cannot work. Even a deflationary move – from plausibility to truth (saying "what's ontologically plausible is true") would be better than having *no* truth here. (We can then debate about whether the deflationary move is justified etc.)

Sure, a theory or principle that's ontologically *not* plausible definitely renders ontology un-understandable. But one that *is* ontologically plausible does not necessarily make ontology understandable (even if it's fairly inter-subjectively objective) if it's not true: it makes the alternate ontology – the ontology of the theory – understandable. May be there is a way to understand the real ontology *via* a false, theoretical one. But if that is the aim then we have to establish clearly some relation between the false one – which is to act as a vehicle of understanding – and the real one, the target object of understanding, and also clearly spell out the way in which this gives us understanding. For instance, if we said that the relationship between the vehicle and target is one of possibility – if the vehicle tells us a possible way in which the target could be or behave, we could get modal understanding as we saw in Chapter 2. The plausibility account of understanding *by itself* can only be a way of understanding the theory, not the world: that a proposed ontology is plausible means we can understand the proposed ontology. But to provide understanding of the world it needs to couple with *some* conception of how the plausible ontology relates to the real world: perhaps it is a correct ontology, or a possible ontology, or provides understanding in one of the other ways discussed in Chapter 2.

If the plausibility account of understanding is to be sustainable by itself, without coupling with any other notions of understanding, it has to involve some kind of truth. But what about Chang's plea to steer clear of truth? According to him, empirical success does not guarantee truth, and neither do ontological principles. His idea is that we needn't be involved in the business of "truth-gambling" at all, since truth is elusive. As he notes, even if we were to come to possess truths about the world, we might not know it. But as I've argued above, the plausibility account of understanding requires truth – some working notion of it, at least. (What conception of truth we should adopt is a separate question, which I won't get into here.) For otherwise, there would be no way of making any claims at all about us understanding the real ontology of our world. So if we don't want to invoke truth at all in the story, the plausibility account of understanding is not sustainable.

While I agree that truth is a problematic concept and that even if we were to come to possess the truth we might not know it; at the least, we can say that empirical adequacy is necessary to get to (at least some approximate) truth (though it of course does not guarantee truth). So even if we remove truth as a goal, for the plausibility account of understanding to hold water, it should at least take into account empirical adequacy: if an ontologically plausible theory is not empirically adequate, then it doesn't meet a truth requirement and hence has no chance of being true. This in turn means that it cannot give us understanding according to the plausibility account of understanding. (It might of course give us understanding on some other counts. In the next chapter I discuss precisely this: how some plausibility-based principles can give us understanding in a different sense without regard to truth/ empirical adequacy concerns.)

So what do we do with a theory that is ontologically plausible but empirically inadequate? As we saw, Chang would perhaps suggest that we use it for the purpose of understanding. But as I've argued, such a theory won't get us understanding of the (real) ontology we want to understand according to the plausibility account of understanding. But Chang is absolutely right in asserting that there is no need to be "faddish" about empirically successful theories and blindly give them our blanket acceptance: it is always good to continue working to find a theory that is both empirically adequate and ontologically plausible. (Chang cites David Bohm as an example of a scientist who cared about ontological plausibility – Chang notes that Bohm followed Einstein in "seeking a version of quantum mechanics without such perceived ontological absurdities as the lack of value-definiteness." (2001, 9)) But again as Chang says, what goals to pursue should be up to different epistemic communities. If one simply aims at empirical adequacy and another at both empirical adequacy and ontological plausibility for the sake of understanding, there's nothing in these goals that makes either superior to the other: both should be acceptable. Here I am with Chang in being voluntarist about aims of science.

Although I think Chang's account of understanding is not sustainable without truth/ empirical adequacy, Chang's views on ontological principles and understanding are certainly not without any merit. Ontological principles – or rather, rational, intuitive principles – like those of single value and no miracles can give us understanding in a different way, by *not claiming to be ontological*, but rather pragmatic. They can give us understanding by helping us intelligibly carry out some of our epistemic activities – and for this kind of understanding, ontological principles don't have to be true. This will be the topic of the next chapter.

3.3 Conclusion

According to the plausibility account of understanding, we understand the world if it is ontologically plausible to us. But a theory or claim that makes ontology come out to be plausible, but is empirically inadequate, does not give us understanding of the ontology we want to understand – it gives us understanding of the ontology that is an artifact of the theory or claim. It is also not desirable to have ontological plausibility as a desideratum for theories – so that our *theories* are understandable – regardless of the nature of reality itself, for plausibility considerations are often subjective and idiosyncratic. Nevertheless, it is undoubtedly desirable to strive for ontological plausibility *alongside* empirical adequacy. It is certainly good to have theories that are both ontologically plausible and empirically adequate. So far, I have argued for the separation of intelligibility and plausibility in Chang's scheme, and that contrary to Chang's view, the plausibility account of understanding is not tenable without having truth/ empirical adequacy in the picture. This concludes my summary and critique of Chang's first paper. I now move on to his later paper.

References

- Chang, Hasok (2001) 'How to take Realism beyond Foot-Stamping' *Philosophy* Vol 76 No. 295 Cambridge University Press 5–30
- Giere, Ronald (2006) *Scientific Perspectivism* University of Chicago Press
- Douglas, Heather (2006) 'Norms for Values in Scientific Belief Acceptance' Philosophy of Science Assoc. 20th Biennial Mtg PSA Contributed Papers
- Douglas, Heather (2012) 'The Value of Cognitive Values' Philosophy of Science Assoc. 23rd Biennial Mtg PSA Contributed Papers
- Laudan, Larry (2004) 'The Epistemic, the Cognitive and the Social' Machamer, P. and Wolters, G eds., *Science, Values, and Objectivity* University of Pittsburgh Press, Pittsburgh, and Universitatsverlag, Konstanz 14-23

4. Ontological Principles and Epistemic Activities

4.1 Overview

In his 2009 paper titled ‘Ontological Principles and the Intelligibility of Epistemic Activities’, Chang advances over the earlier work and steers away from realism as the subject matter. Here his position on ontological principles is a little different: they are taken to be *relativized*: they are relative to the “epistemic activity” at hand. Going back to the earlier example, the present version would be: *if* for instance we are to intelligibly engage in the activity of testing a theory by comparing its prediction of the value of a physical property with the value yielded by experiment, *then* we *have to* assume the principle of single value. If we didn't assume it, the *activity* would become incomprehensible and senseless. The principle is not a metaphysical truth that holds unconditionally: its roots are entirely pragmatic. *We need* to consider it to hold *if we* are engaging in an activity whose very intelligibility relies on the principle. Otherwise, we need not be committed to it. (Chang also expresses very similar views in his 2008 paper on Kant and modern philosophy of science titled ‘Contingent Transcendental Arguments for Metaphysical Principles’.)

Here I critically evaluate and improve upon Chang’s account. I point out that intelligibility is just one facet of an activity we can demand to have understanding – there can be others like practical utility and an ultimate meaning or purpose to carrying out the activity. I then argue in favor of Chang’s view of this pragmatist account of understanding: understanding consists in carrying out a pragmatically chosen epistemic activity intelligibly/ meaningfully. How does this give us understanding of the *world*? Straightforwardly, it doesn't. Rather, this kind of understanding involves *us taking ontology to be a certain way* – conforming to ontological principles – so that we are able to carry out our desired activity fruitfully. But as Chang points out, we only interact with ontology *via* our epistemic activities, so arguably this understanding we get is not just of our activities, but of the world as well – the world that *we* encounter and interact with.

4.2 Ontological Principles and Epistemic Activities

4.2.1 Background and Exegesis

The central idea in this work of Chang's is that the necessity of ontological principles are contingent upon our epistemic[11] activities – activities *we choose t*o engage in to fulfill whatever pragmatic needs we may have – and thus are entirely pragmatically rooted. So if for instance we didn't have to engage in an epistemic activity that involved comparing two or more values of the same physical property like temperature obtained from different methods; we wouldn't have to be committed to the principle of single value. Before I begin I'd like to point out that Chang does not once use the word 'plausibility' in this paper: his arguments are now entirely in terms of intelligibility.

---

[11] By 'epistemic activity' Chang does not mean an activity that proceeds by making true assumptions or that yields truth or knowledge. Chang is not committed to truth at all and uses the term much more broadly as we shall see in his examples. So he probably shouldn't have used the word 'epistemic', but since this is a report and analysis of his work, I will use his term as he does.

Further, ‘intelligibility’ is now used for activities rather than for principles or claims: an activity is intelligible or unintelligible depending on whether we have presumed the relevant and required principles.

Chang begins with a discussion of Leibniz' principle of continuity. Leibniz attacked Descartes's second rule of two colliding bodies which said that if both bodies approached each other with the same speed and one were larger than the other, then only the smaller one would rebound and then the two would continue in the same direction with the larger having more force than the smaller; and the larger would not be forced to rebound by the smaller. Now it might seem obvious that Descartes was wrong about this but the crucial point for Chang is that Leibniz did not resort to experiment to refute this rule. For Leibniz, this rule was a “metaphysical absurdity” for it violated an ontological principle: a principle of continuity according to which if a cause varies continuously, the effect cannot vary discontinuously. Chang draws inspiration from this episode to make a case for non-empirical factors playing a significant role in our knowledge systems. He says,

> The contrast to standard empiricist schemes could not be clearer: in Bas van Fraassen's (1980, 87-96) view, for example, something like intelligibility would probably be treated as a "pragmatic virtue", a merely subsidiary criterion that might function as a tie-breaker if a choice were desired between two theories of equal empirical adequacy; with Leibniz, it is clear that intelligibility was just as important as empirical adequacy, if not more important. What is also nice about Leibniz's thinking is that the verdict of unintelligibility is based on a clearly articulated ontological principle that seems beyond dispute at first glance. This is rationalist physics at its best. (2009, 66)

However, in contrast to what Chang says about this principle in his earlier paper (where he takes it as quite a serious candidate for an ontological principle), in this paper, he takes it to be merely a "ladder" to understanding intelligibility as a virtue, but only to be "kicked away".

Why? Because according to Chang, while the rationalist motivation behind it can be appreciated, it is subjective, idiosyncratic, and uncertain. Chang repeats many of the historical examples (Einstein's distaste for probability in quantum mechanics; Cartesians' disbelief in ian action at a distance etc.) from the earlier paper; but contends that while we should learn from these the spirit of not blindly yielding to empiricism, we need firmer ground for establishing ontological principles. And for him, this can be achieved by embracing only those beliefs whose denial is unintelligible – what he calls ontological *principles* – as opposed to ontological *opinions*, which are like Liebniz's belief discussed above: their denial may be "strange" but not strictly unintelligible and hence we cannot be certain about them. (His idea of intelligibility is finer than this: strictly speaking it is not the denial of the principle that is itself unintelligible, but the activity whose associated ontological principles are being denied. I will get to this in a bit.) Importantly, denial of ontological principles strikes us as nonsensical rather than false – they provide a basis for *intelligibility* and *not truth.*

Chang then goes on to analyze the nature of the necessity of ontological principles. The initial arguments are similar to those in the previous paper. Two natural guesses, empirical and logical necessity, are ruled out. It would make no sense to test these principles empirically: measuring the temperature of something multiple times to check if we get only one single value each time he says is akin to testing "All bachelors are unmarried" by collecting a number of bachelors and checking carefully to see if each of them is married. (This goes back to his point made in the earlier paper about ontological principles being untestable). Secondly, he argues, these principles are not logically necessary either: it is not a logical contradiction to assert that a physical property has two values at once.

Chang concludes that the source of this necessity lies in *our pragmatic requirements.* They are similar to Kant's synthetic a priori, but unlike Kant's a priori principles, Chang's principles are not eternal and universal: they are entirely rooted in our pragmatic goals and pursuits and can be conveniently and unproblematically disregarded when our activities are such that they do not require them to hold. In the case of the principle of single value, the (pragmatic) conditionality can be expressed as follows: If (and only if) we want to engage in an epistemic activity, then we have to presume corresponding ontological principle(s). If for instance we want to engage in the epistemic activity of testing a theoretical prediction of a physical property with the experimental prediction (among other similar activities like the one Regnault engaged in as discussed in the earlier section), *then* we *have to* subscribe to 'single-valuedness'. Otherwise, we can suspend our commitment to it. (Whether we ought to engage in a particular activity is a separate question, he says.) Similarly, if we want to engage in the activity of rational prediction, we have to subscribe to the belief that the same initial condition will always have the same final outcome (induction). Chang calls this belief the "principle of uniform consequence". Another example is that of counting: if we want to engage in the activity of counting, then we have to subscribe to the view that the things being counted are discrete: "the principle of discreteness". This way, Chang moves a step forward from his earlier paper and offers a way to make sense of how ontological principles can be necessary yet not fixed and universal but rooted our pragmatic interests.

To further highlight this point, he says:

> Necessity also does not imply universality. Ontological principles are not truly universal, as their necessity remains conditional on our commitment to engage in their corresponding epistemic activities. We may not see this usually, but there are situations in which we would decline to engage in certain basic epistemic activities, such as rational prediction, or counting, or explanation. (2009, 74)

Of course, Chang does not mean here something as trivial as that the necessity of ontological principles operates via some on-off switch: turning on when we engage in the relevant activities and turning off when we don't – that would be absurd. What Chang means, I think, is that if we believe in engaging in an activity in a particular situation – i.e. we think engaging in it is acceptable, that it is possible to get acceptable results by engaging in it, then we cannot deny the corresponding ontological principle irrespective of whether we actually engage in the activity – hence the word "commitment". If for instance, we don't believe in engaging in rational prediction in a particular situation for we don't find it acceptable – say we resort to divination, then the principle of uniform consequence is not necessary for us. Since some activities like counting are so basic and pervasive, the corresponding principles *seem* to be unconditionally necessary: we are *always* committed to the activity of counting in that we always find it rationally acceptable (that it may be irrelevant in some situation is beside the point), so we will *always* find the principle of discreteness necessary. We should not let such cases mislead us. Chang's conditionality still very much holds. If counting were something we didn't always find rationally acceptable to engage in, or further still if we had never come up with the idea of counting, we would have never found the principle of discreteness necessary (this is reminiscent of C.I. Lewis' allusion to the lack of arithmetic in a jellyfish world (Lewis, 1929, 252)). On the other hand to take another of Chang's examples, say in a particular situation we think it doesn't make sense to 'engage in the activity of empathy' – empathize – then we no longer find it necessary to believe in the existence of other minds.

Importantly for Chang, an ontological principle, probably contrary to what the term might suggest, does not state that something exists:

> Rather, the term "ontological" indicates a concern with the basic nature of the entities that we are thinking about. The entities in question may be of any type (for example, physical or abstract), but in any case they will have some basic properties and the ontological principles dictate what kind of properties they can have. (2009, 69)

It is worth noting that while in the earlier paper Chang considered statements (of ontological principles) to be unintelligible if denied, here he considers primarily the activities, and only secondarily statements, propositions etc. to be unintelligible if ontological principles are denied. Before moving on, here is an important clarification.

Although Chang patently asserts that it would be *impossible* to engage in an activity without assuming the relevant ontological principles (70; also 75: "[we] would not be able to perform an unintelligible epistemic activity"), I don't think this is what he exactly means: as he says in other parts (2009, 70) it would be *unintelligible* to engage in the activity without assuming the relevant principles and I think he uses "impossibility" as a shorthand for "impossibility of intelligibly performing".

Through his scheme of epistemic activity and ontological principle pairings, Chang envisages a deep and close link between ontology and epistemology:

> …in the principle-activity pairing, the ontological principle and the epistemic activity are mutually constitutive. A general link between ontology and epistemology will be easily granted, in the sense that the appropriate method of studying something is surely linked with the nature of that something. Here I am pointing to the purest and strongest version of that link: a distinct type of epistemic activity has its particular type of object, the essential characteristic of which is defined by the fundamental ontological principle associated with the activity. At the same time, the nature of the activity is shaped by the basic character of its objects. (2009, 71)

Below are some of what I found to be interesting epistemic activity – ontological principle pairs:

- Counting – discreteness: "It is only if we want to count things that discreteness has to be presumed" (2009, 70). Note the "only" clause: so for Chang, it is *if and only if* we are committed to engaging in an activity that the relevant ontological principles need to be presumed.

- Rational prediction – principle of uniform consequence or induction: In order to make a rational prediction (as opposed to lucky guess or divine intervention), we need to assume that the same circumstances will always have the same outcome (2009, 71). Otherwise, our activity of rational prediction becomes unintelligible.
- Narration – principle of subsistence: In Chang's words, “A narrative requires entities or persons whose identities last through time, which "house" the changes that are narrated. Otherwise, we cannot even formulate narrative strands like "someone ran down the street", or "on earth, first this happened and then that happened". Paradoxically, without postulating <u>something</u> that lasts, it is impossible to describe any change. “ (2009, 73)
- Linear ordering – principle of transitivity: Again, if we deny transitivity of our method of ordering, then any attempt to put any kind of things in an ordered sequence will be unintelligible and fail.

Highlighting the pragmatic nature of these links, Chang makes another (related) distinction between ontological opinions and ontological principles. The former have no pragmatic consequences – for denying/ abandoning them doesn't make it impossible to intelligibly perform any pragmatically important activity – and therefore may be discarded; while that is not the case with the latter. Necessity, according to Chang is entirely rooted in human activity. So it seems like he's saying that principles that simply seem necessary without being necessary *for* an activity, are suspect because they are (pragmatically) baseless.

To summarize the above in the form of a standard (deductive) argument:

- Premise 1: If we want the epistemic activities we find acceptable and want to pursue to be intelligible, then we cannot deny the relevant ontological principles.
- Premise 2: (As rational beings) we do want the epistemic activities we pursue to be intelligible.
- Conclusion: We cannot deny the relevant ontological principles.

Further, he says the success of our activities would serve as a loose vindication of the corresponding principles – and this might hint at some kind of realism. However, Chang is wary about carrying this too far: "But this vague sense of vindication is all we can have, and it only points to an inarticulable harmony between the state of the world and our ontological principles" (2009, 74), he says. One might sense a borderline antirealist, or an instrumentalist, streak in Chang when he says,

> When nature "speaks" to us, it is only through the outcomes of our epistemic activities. That is not to deny that nature enters and rules our life, by determining which epistemic activities are pragmatically possible. But ontological principles do not give us the kind of direct representation of nature that the correspondence theory of truth would prompt us to seek. (2009, 75)

Chang thinks that ontological principles do not give any "direct" representation of (parts of) the world. *We* tend to associate success with truth and since we are not very sure of the legitimacy of this association, the vindication leading to truth is but a vague one. The nature and strength of the vindication Chang discusses is however not very clear: going by his view that the vindication indicates a kind of a "harmony" between our ontological principles and the world, he seems to think that our ontological principles are (at least roughly) consistent with some underlying metaphysical truth – but how to make sense of this connection between success of an activity and truth of presumed principles is not very clear. I will take up a discussion of this in the next section, but for now, a moral analogy that might help: Say a parent has no choice but to deceive her child about the existence of a monster and instill fear in order to make her eat her food. If the parent is successful (success could be measured by the child eating well, growing up to be healthy etc.), then she could assert that deceiving the child in that situation is morally permissible and that the moral principle "Deceive your child to make her eat well" is *vindicated* – for it was pragmatically necessary for the purpose of engaging in the activity of feeding the child – and therefore morally correct and indicates a "harmony" between our actions and moral truths.

Once again, as with the previous paper, Chang attempts to establish a close link between intelligibility and understanding. He begins by equating intelligibility of an activity to its performability: "intelligibility is the performability of an epistemic activity" (2009, 75). As mentioned briefly earlier, this is not very convincing for it seems that an activity can possibly be performed even if it is unintelligible (only, we might be irrational in engaging in such an activity). But for Chang "performing" seems to automatically mean "intelligible performing" which is why he equates intelligibility to performability. He admits that this is a minimal conception of intelligibility but says it is not an entirely vacuous one, for "Performability requires a certain harmony within the activity. For example, any statement made within it needs to conform to the ontological principle associated with it." (2009, 75) So where does understanding figure here? Keeping with the pragmatist conception of intelligibility, Chang has a pragmatist conception of understanding too: "Understanding is knowing how to perform an epistemic activity" (2009, 75) He goes on to say,

> Understanding, as I see it, is not some distinct quality or state of mind that exists separately from the sense of knowing how to do epistemic things. Understanding is simply knowledge taken in the active sense. The feeling of understanding is the sense of performing an epistemic activity well: "I know what I am doing." The degree of understanding is the degree of success with which we engage in an epistemic activity, as assessed by ourselves or by others. (2009, 76)

For Chang, the difference between mechanically solving a problem applying some rules of arithmetic and having an understanding in solving a problem lies in correctly recognizing what epistemic activity we are actually interested in and engaging in; and then applying the "right tricks" to successfully carry out the activity: "Any well-performed epistemic activity can generate a sense of understanding."(2009, 76) As an aside, what "well-performed" means is not explicated by Chang. While not denying relevant ontological principles is a necessary part of performing an activity well, it is not clear what else goes into a well-performed activity: I return to this in the next section where I discuss success of an activity. Since people have different epistemic aims, what they take to be understanding will also vary with the aims. For Chang, this pragmatic subjectivity is deep rooted and ineliminable in our epistemic goals and therefore also in what we take understanding to be.

In concluding this paper, Chang gives a second list of epistemic activity – ontological principle candidate pairs of which he is less sure. To list a few:

- Empathy – other minds: if we want to engage in empathy, we have to subscribe to the belief that there are minds other than our own
- Deductive inference – noncontradiction: if we want to engage in a deductive inference, we have to subscribe to the belief that there cannot be a contradiction. (This is strikingly reminiscent of Herbert Feigl's (1950) views in his 'De Principiis Non Disputandum...' According to Feigl, the principles of logic cannot be given justification in the sense of validation; their "ineluctable character" is accounted for by pragmatic vindication: "If pragmatic *vindication* is sharply distinguished from validation, then all it can provide amounts to a recommendation of a certain type of behavior with respect to certain ends. We may say to ourselves: If we wish to avoid the perplexities and discomforts that arise out of ambiguity and inconsistency, *then* we have to comply with the rules of semantics and logic. If we wish to derive true propositions from true premises then we must conform to the rules of inference and the rules of substitution. The reasoning concerning these means-ends relations utilizes, as any such reasoning must, the forms of deductive and inductive inference." (emphasis as in original))
- Observation – externality/ objectivity: if we want to engage in observation, we have to subscribe to the belief that the things being observed have an objective (i.e. extra-subjective) externality.

He acknowledges that his list (the earlier and this one put together) only consists of simple activities and that most everyday, real activities are complex ones, often conjunctions of many simple activities. For these to be intelligible and performable then, by extension, each of the ontological principles corresponding to each of the constituent activities respectively have to all be satisfied.

Chang closes with an important distinction between his conception of understanding in the earlier paper and the present one: while in the earlier one he treated understanding as a goal separate from prediction and description, here, he takes it to be at the heart of just any well performed activity – including prediction and description. Therefore understanding becomes part of all the aims of science and all epistemic activities in general. And finally, he reiterates the point that he is not concerned with truth-finding here, but only wants to establish intelligibility as an epistemic virtue (recall, "epistemic" in his sense – unrelated to truth), independent of truth-finding. His view is that we could indirectly arrive at some kind of truth, in the sense of "harmony with the world" – the truth of at least some of our ontological principles – via inference from the successful performance of the respective epistemic activities.

4.2.2 Reflection and critique

I want to begin by clarifying an ambiguity in Chang's narrative of ontological principles. What exactly is it that we do with the relevant ontological principles if we want to intelligibly perform an activity? Chang is not consistent with choice of words here. He uses “subscribe to them”, “assume them”, “presume them” and “not deny them” all interchangeably. I would like to point out that all these phrases are not exactly equivalent in meaning. We don't always actively/ consciously “subscribe to” or “assume” the principles. If this were the requirement for understanding, then we could not claim to have an understanding of most of the things we do, for we seldom actively *assume* that the things we count are discrete or that our linear ordering of things is transitive. Hence to say that we “presume” the principles makes more sense: there is an inherent tacitness in the word “presume”. Even better is to say we *don’t deny* the principles – for since presumption is tacit, the only way of testing whether a principle is presumed is by asserting its denial and checking if that denial results in some kind of a disagreement within us.

Now getting into a critical analysis of Chang’s work, I want to start with the point that like theories, activities can also be expected to have values, and the familiar questions about (theoretical) value choice can arise here as well. For instance, simplicity may be preferred as a value for the method of carrying out some activity. For Chang, the activity needs to have the value of being intelligible. As before, I contend that Douglas' distinction of criteria vs. values applies here as well and that intelligibility is again, in Douglas’ terms, a baseline criterion, not a value per se. i.e., we always want our activities to be intelligible at the least. Claiming to count while at the same time denying that the objects being counted are discrete would be *inconsistent* with respect to the activity. I mean, to simultaneously assert that I am counting objects and that they are not discrete would be inconsistent. So in this case, intelligibility lines up with an already established criterion: internal consistency (here, internal to the activity). As Chang himself says, an epistemic activity and the relevant ontological principle are mutually constitutive (in fact in the 2008 paper he goes so far as to say that a statement like "Counting requires discreteness" is analytic). This being the case, when an activity is performed, the accompanying ontological principle is to be *necessarily* presumed. It is a "baseline criterion" that is required to prevent inconsistencies, rather than a virtue or value. So now that we have taken intelligibility to be a basic criterion for epistemic activities, what kind of values can we expect our activities to have? We can expect our activities to have as values simplicity, broadness of scope – can the activity be performed in a wide range of situations/ to address to a wide range of problems? – usefulness, purpose or justification – is engaging in the activity purposeful/ pragmatically justified?, etc.

With activities for which we want these other values, the conditionality of ontological principles is probably not as rigid as it seems in the case of intelligibility. With the last of the values I listed above for instance, the principle might not render the activity unintelligible, but might preclude us from meaningfully engaging in/ performing an activity in the sense that we don't find a purpose or adequate justification for carrying out the activity in the first place. So the necessity of the principle is still pragmatically rooted, but is probably not as strong as in the case of intelligibility.

To take the Hacking scenario from Chang's earlier paper for instance, what happens when we deny the principle of cosmic honesty if we are committed to the activity of studying the microscopic nature of some object? The activity is not rendered unintelligible in the way say, the activity of counting is rendered unintelligible by denying the principle of discreteness. But if what we observe through multiple microscopes all agree with each other and yet we deny that they all correspond to the real structure of the object in question, then our activity of studying the microscopic structure of the object becomes meaningless and futile: if we are committed to studying the structure of the object, we are also (tacitly) committed to the idea that multiple distinct and properly functioning microscopes, if all reveal the same structure of the object being studied, what we're observing through each is in fact the true structure of the cell. In the previous paper, Chang says about the possibility that all the microscope images being the same is purely coincidental, "It is extremely hard to believe that nature would behave in such a way as to deceive us intentionally, and with the same result time and again." (2001, 28) There, Chang's point was that this kind of deception is implausible. But here I want to argue that there is a coherent story to be told about this situation, that fits with Chang's current activity-and-principle scheme. Why do we find the denial of the principle of cosmic honesty implausible? What's at stake if we deny it? Not intelligibility, but I think it is a meaningful end or purpose – a justification for engaging in the activity. I contend that when we are committed (in the sense explained earlier) to the activity of studying the microscopic structure of the object, the activity will become purposeless – a wild goose chase – if we deny that what we're observing under several different microscopes is indeed the structure of the object. It will be a pointless exercise to engage in the activity if we deny that we are meeting the purpose.

Further, as with intelligibility, here again where we engage in an activity that is without a meaningful end, there is a lack of *understanding* in engaging in the activity. For having a proper understanding, we cannot deny principles to which the very purpose of pursuing the activity is tied. We can probably call this kind of understanding a teleological understanding of activities – understanding got from engaging in an activity with a fruitful end or purpose. (What counts as fruitful of course, is again to be pragmatically decided by the relevant community.) It should be clear by now that denying relevant ontological principles can rob us of our understanding of the corresponding activities: if the principles are denied, it no longer makes sense to engage in the activity – to take two of the examples of values for activities discussed so far, denying ontological principles can render activities a) unintelligible or b) teleologically unjustified. But all this has to do with performing the *activity* with understanding. How does this relate to understanding the *world*?

The simple answer I think is that it doesn't, in a straightforward way – i.e. in a realist sense. Arguments similar to the one I advanced against plausibility giving us understanding of reality in the previous chapter can be made here: ontological principles presumed to make a corresponding activity intelligible or justified do just that: make the activity intelligible or justified, or in a word, understandable. There is no guarantee that they tell us anything about ontology itself. But we can give a *pragmatist* account of understanding here (in some ways similar to de Regt's account discussed in Chapter 2) according to which the principles need not reflect features of reality itself. In this account, understanding simply consists in, as Chang says, performing an activity well – meaning (possibly among other things), the relevant principles are presumed and not denied. Unlike the plausibility account of understanding, here the claim is not that ontology has to have certain plausible/ intelligible features in order to be understandable. It is rather that presuming relevant principles simply helps us get on with our business. Our pragmatically chosen activities become intelligible and meaningful. The crux then, is that we understand reality in a way that is pragmatically prudential for us. One may ask how this truly gives us understanding of objective reality, outside of *our human* pragmatic concerns. Again, the answer is that it doesn't give us an objective understanding in this sense, but such an objective understanding is not even the goal here. But it is also important to note as Chang says, that we only interact with and know about reality *through* our epistemic activities – we don't have a way out of that. So it could be argued that it is a feature of reality to conform to our pragmatic goals in the sense that it behaves in a way that enables our intelligibly engaging in our chosen activities. For instance, as far as our activity of testing a theory by comparing theoretical and experimental values of a physical quantity, nature behaves (to us) as though the principle of single value were true. This I think, is what Chang means by the "inarticulable harmony" between our ontological principles and the state of the world.

What about empirical adequacy in this story? If a principle is found to be empirically inadequate do we still hold on to it and not deny it if it helps intelligibly/ purposefully perform an activity? That would depend on the situation – on the tradeoff between empirical adequacy and the pragmatic fruitfulness of the activity. If there is a high pragmatic payoff by engaging in the activity – to be decided by the relevant epistemic communities – we can still hold on to the principle for the sake of engaging in the activity. In the earlier story in the previous chapter, I said that if a theory is not empirically adequate, but ontologically plausible, it does not give us understanding of the ontology we want to understand per the plausibility account of understanding. Here, the goal is a pragmatic one – to intelligibly/ meaningfully carry out activities. So I don't see a problem in adopting a principle that advances the goal even if it is not empirically adequate.

Can we spin a similar story about the plausibility account of understanding? Can we say that just like we want our activities to be intelligible without being concerned with the nature of reality, we hold on to plausibility principles just so our theories and claims are plausible, and not care about the 'real' ontology? The problem with this approach is that neither does it care about the nature of reality, nor about any pragmatic goals. But if we can give good reasons for pursuing ontological plausibility – for its own sake or for other specific purposes – then sure, we could strive for theories that are ontologically plausible. It seems that principles that clearly help perform activities with understanding enjoy more consensus, but there could be plausibility-based principles that aren't pragmatically rooted in this way but still enjoy reasonable consensus. But I'm doubtful if there could be good, objective reasons for pursuing ontological plausibility as a goal.

For those interested in truth and epistemic merit, it seems that pragmatically rooted principles fare better. It doesn't seem very plausible that principles not rooted in any concrete, pragmatic activity, even if shared across many people have any good shot at truth. With principles that are pragmatically rooted, success is a differentiator: as Chang says (as Duhem, I argue, in Chapter 6), when we succeed at a chosen activity (say counting or prediction) our faith in the presumed principles (discreteness, uniform consequence respectively) is reassured. Sure, this might just be 'faith' (for lack of a better word), but we at least get this "vague sense of vindication" for the principles as Chang says. But with plausibility-based principles that serve no pragmatic purpose, we don't have even this vague sense of vindication: since there are no mutually agreed upon activities with goals, there seems to be no objective way to talk about success. Hence there seems to be no meaningful and objective way to talk about these principles being vindicated or saying anything about reality.

All the *intelligibility* (as opposed to plausibility) principles from the earlier work discussed in Chapter 3 that seem to have unequivocal, universal appeal and don't seem to be subjective and idiosyncratic – like the principle single value – owe their (inter-subjective) objectivity to the universal *activities* they are rooted in. (For the principle of single value the activity would be that of testing a theoretical value of say temperature with an experimental value as discussed earlier.) It might seem like these principles are unconditionally necessary to us, but as explained earlier, we are always committed to the respective activities (even when not performing them), so we never deny these principles. That doesn't make them unconditional: they are still very much pragmatically rooted. Other, more shaky principles like that of contact-action advanced by the anti-Newtonians or Einstein's dictum that God doesn't play dice – don't seem to be (inter-subjectively) objective and don't enjoy universal appeal since they are not pragmatically necessary –

there don’t seem to be specific pragmatically chosen activities that they help epistemic communities intelligibly/ meaningfully carry out. They are simply individual and subjective idiosyncrasies that neither latch on to reality nor help achieve our pragmatic ends – they are, in Chang’s terms, “ontological opinions”.

Turning now to Chang’s realist step, he says that success of our activities indirectly vindicates the ontological principles involved. Since he goes on to say that this points to an “inarticulable harmony between the state of world and our ontological principles”, I take it that by vindication he means (we have some vague assurance that) we're getting something at least roughly right about the world: our ontological principles needed for intelligible performability of our activities are somehow aligned with the state of the world. Several things need to be clarified here. What does success consist in? Let us say it is success normally construed: however the relevant epistemic community defines it, according to its needs and goals. For instance, we succeed at the activity of counting when we do it intelligibly and get a result that is deemed correct, meaningful, and acceptable. Chang also relates understanding to success when he says that a well-performed (which I take to mean the same as “successful”) activity generates understanding. Further as we have seen, intelligibility is key to understanding for Chang. But he admits that while it is necessary, it is not sufficient. Putting these points together, understanding is measured by the successful performance of an intelligible activity. In other words, we gain understanding when we perform an intelligible activity, and the activity yields good results/ furthers some pragmatic goal etc.: basically, understanding consists in intelligible performance of an activity plus success as decided by the relevant epistemic community.

How to make sense of this connection that Chang tries to make between success of activities and vindication of principles[12]? One way to relate them may be to make a holist connection: like in confirmation holism where confirmation of a theory confirms all the assumptions made in the theory, here success of an activity vindicates all assumptions required for intelligible performability of the activity – and therefore ontological principles. Of course, the comparison only goes so far. To begin with, there is no confirmation here, and the relationship between a theory and its assumptions is certainly not the same as that between an activity and principles assumed for its intelligibility. Nevertheless, such a holist move seems to be the only meaningful way to make sense of the connection that Chang tries to make between activities and the assumed principles. Although it won't be confirmation holism, we can try and come up with a different kind of holism, say "success-vindication holism". Loosely, the idea is that when an activity is successful, the success should be attributed to all components of the activity including presumed principles that make the activity intelligible, purposeful etc. Of course, the step from success to vindication of the principles has to be based on some kind of ampliative reasoning as in the realism arguments. This could take various forms: we could say it is implausible/ unintuitive/ against any reasoned guess, that given

[12] In a footnote, Chang likens it to Duhem's move of inferring from the progress of physical theories that they are converging onto a metaphysical reality. I don't think this is an appropriate comparison since here the premise and the conclusion are about the same thing: physical theory; whereas with Chang, the inference is from (success of) *activities* to (vindication of) *principles* and is hence a little more complicated.

the success the presumed principles are false (or that it is plausible/ reasonable/ rational to take the principles to be vindicated based on success of the activities). It is for this reason that Chang says we can only have a "vague sense" of the vindication.

4.3 Conclusion

In the pragmatist account of understanding, we want our chosen epistemic activities to be intelligible and meaningful to us. To this end, we have to presume and not deny certain ontological principles. Understanding here consists in carrying out these activities intelligibly, purposefully, successfully, etc. Ontology is taken to be in a way that is pragmatically prudential to us; in a way that gives us understanding of our epistemic activities – and if we succeed in our activities, our ontological principles are indirectly, vaguely vindicated.

References

- Chang, Hasok (2001) 'How to take Realism beyond Foot-Stamping' *Philosophy* Vol. 76, No. 295 Cambridge University Press 5–30

- Chang, Hasok (2008) 'Contingent Transcendental Arguments for Metaphysical Principles' in Michela Massimi, ed., *Kant and Philosophy of Science* Cambridge University Press 113-134

- Chang, Hasok (2009) 'Ontological Principles and the Intelligibility of Epistemic Activities', in Henk de Regt, Sabina Leonelli, and Kai Eigner, eds., *Scientific Understanding: Philosophical Perspectives* Pittsburgh: University of Pittsburgh Press 64–82

- Feigl, Herbert (1950) ‘De Principiis Non Disputandum ...’ in Max Black, ed., *Philosophical Analysis,* Ithaca: Cornell University Press Retrieved from http://www.ditext.com/feigl/pnd.html 05/26/2015

- Giere, Ronald (2006) *Scientific Perspectivism* University of Chicago Press

Part II – Understanding of Science

5. Philosophy of Science in the time of Pierre Duhem

5.1 Introduction

Pierre Duhem was a late 19th – early 20th century physicist, philosopher, and historian of science, and a contemporary of Karl Pearson, Henri Poincaré, and Ernst Mach. In physics, he worked primarily in thermodynamics and made important contributions to the area of energetics. In philosophy of science, he is probably best known for his work on the relation between theory and experiment, particularly his view of underdetermination of theory by evidence – the idea that hypotheses are not singly refuted by experiment and that there are no crucial experiments in science. In history of science, he produced groundbreaking work in medieval science and defended a thesis of continuity between medieval and early modern science.

In this chapter, I give a historical backdrop of science and philosophy of science during the time of Duhem, and discuss – for the sake of useful comparison – the views of Pearson, Poincaré, and Mach. In the next chapter I take up a detailed study of Duhem's philosophy where I draw parallels between Duhem's way of making sense of and understanding the practice of science, with Chang's account of making sense of and understanding the world in our science. Similar to Chang, I show that Duhem argued that a physicist had to make a (realist) presumption that was metaphysical in nature, in order to make sense of her practice of coming up with theories.

Interestingly as we'll see, all of these people have been instrumentalist/ antirealist (in probably slightly different senses of the terms) about theories in science, and all of them were in some sense against metaphysics and wanted to guard science from it. Historically, this seems to have been a period when scientific and metaphysical antirealism were strong: the general view was that scientific theories do not get to some underlying, metaphysical truths, and further, that an external world doesn't exist was held by many distinguished thinkers. But I believe Duhem was an exception: while a superficial reading of him might lead one to interpret him as being instrumentalist and anti-metaphysics, a closer reading reveals that he in fact had a quite realist attitude towards physical theory, and thought that a thinking, contemplating scientist cannot do without metaphysics: she needs metaphysics to make sense of her own practice of science. So I think Pearson, Mach, and Poincare – who were quite clearly instrumentalist about theories in science and were against metaphysics – together provide a nice contrast class to study Duhem: despite many apparent similarities, there are deep differences.

I contend that Pearson, Mach, and Poincare were not so interested in *understanding* the world and science the way Chang and Duhem are: they were interested in having epistemically good and correct beliefs about the world, as well as about science. On the other hand as we saw in the last chapter, one can meaningfully have a pragmatic approach towards metaphysical claims about the world: one needn't claim that one *knows* that an external world exists, for instance, but just make the affirmation that it does, for the pragmatic purpose of intelligibly engaging in the activity of observation. Further, as we'll see in Chapter 6 about Duhem, one can have a certain view of science (for Duhem a realist view) for the reason that it is pragmatically useful – if it helps the scientist make sense of her own practice and gives her purpose and direction. But again, I don't think Pearson, Mach, and Poincaré were interested in such understanding of science – they were rather interested in the epistemology of science.

5.2 Science and Philosophy of Science in the time of Pierre Duhem – an Overview

5.2.1 Historical Backdrop – A Brief Account

The late 19th century was an important period in the history of western science – it marked the birth of some seminal ideas including Darwinian evolution, Pearson's relativity of motion which was a precursor to Einstein's special theory of relativity, and Poincaré's dynamical instabilities that paved the way for modern nonlinear dynamics and chaos theory. Some of the scientific implications of these ideas led to some tremendous changes in beliefs and attitudes towards science and the universe it describes. A big chasm developed between science and religion; a lot of thinkers wanted to keep science free of any metaphysics as well, and the "clockwork universe" – the idea that the universe works like a perfect mechanical clock governed by the laws of physics, with every past, present, and future aspect of it perfectly determinable – was in disarray. Here I briefly survey some elements of these developments, followed by overviews of the philosophies of three contemporaries of Duhem who were towering figures in the period: Ernst Mach, Karl Pearson, and Henri Poincaré.

Perhaps one of the most striking features of science in the period was the disengagement with religion. According to Darwin's theory, evolution was marked by random variations and competition for scarce resources – and this completely went against theological arguments from design. Darwin's theory of natural selection implied that the world was far from being well ordered and harmonious, contrary to Biblical thought. While different thinkers of the time might have had varying personal views on religion, the prevailing view on science and its practice seems to have been that scientific practice must be free of the influence of religion as well as metaphysics. For instance, Poincaré is said to have been a believer in his early years, but progressively turned agnostic. (Rollet, 2014, 9) On the other hand, Duhem was a devout Christian but yet held that physics was entirely distinct from metaphysics and that physicists should not, in their scientific practice, concern themselves with metaphysical questions such as that of the true, underlying nature of reality. (This is only part of the story though and I say a bit more about this towards the end.)

Another development that had a profound influence on the way the universe came to be understood was the uprooting of the Newtonian worldview. Darwinism shook the Newtonian vision of regularity in the universe; further, Pearson was already anticipating the revolutionary ideas of Einstein's special theory of relativity. Pearson insisted on the relativity of all motion, and reframed the Newtonian laws of motion to conform to this central idea. And as James Kloppenberg (1986) notes, Mach's *Science of Mechanics* (1883) also contributed to the weakening of Newtonian mechanics: it contained experimental evidence indicating that, "absolute space and motion, two of the mainstays of Newtonian mechanics, are, in Mach's words, no more than pure mental concepts, which "cannot be produced in experience."" (Kloppenberg, 24) Further, the Cartesian model in geometry was dismantled – this was the period when non-Euclidean geometry was gaining prominence. This also meant Kantian the a priori was rejected – it was based on the presupposition that Euclidean geometry and Newton's theory were true, so once these latter two were superseded, the Kantian a priori was no longer tenable.

Despite such new, revolutionary, and persuasive ideas in science and the rift they created with religion and metaphysics, the general view was not that scientific claims were literally true of nature: the last two decades of the 19th century saw the questioning of claims to certainty based on science. For instance, Pearson (1892) argued that concepts within science such as regularity and order do not reflect workings in nature itself, but merely act as a "conceptual shorthand" without addressing metaphysical questions about reality. A major contributing factor to this attitude was the realization of the problem of underdetermination notably by Duhem and Poincaré. As Kyle Stanford (2006) notes,

> The progress of physical science had by this time begun to suggest that there might be quite genuine cases of differences between actual competing scientific theories that could not possibly be adjudicated by any straightforward appeal to empirical tests or observations. To use a famous example of Poincaré's (though not a case of actual competing theories), any set of measurements of the angles in a triangle marked out by appropriately oriented perfectly rigid rods can be accommodated by the assignment of any number of different combinations of underlying spatial geometries and compensating 'congruence relations' for the rods in question; if the sum of the angles differs from 180 degrees, for instance, one may either interpret the underlying geometry as Euclidean and conclude that the distance marked out by each rod varies with its position and/or orientation, or assume that the distance marked out by each rod

> remains constant and conclude that the underlying geometry of the relevant space is non-Euclidean. Poincaré's response to this problem of theoretical underdetermination was conventionalism; that is, he regarded such theoretical matters as the assignment of a particular physical geometry to space as matters of choice or convention to be decided on grounds of greatest convenience. And this in turn implied, he suggested, the distinctively instrumentalist conclusion that the quite useful ascription of a particular geometry to space by a theory should not be construed as literally attributing anything (truly or falsely) to nature itself. (2006, 401)

Instrumentalism towards not just geometry, but science and its theories, was indeed the dominant view at the time and is a striking commonality in the philosophies of Duhem, Poincaré, Mach, and Pearson. All of them insisted that science must be free of metaphysical concerns about the nature of reality, and that all scientific theories are, are tools that organize and classify empirical laws and phenomena for the sake of convenience. But it is important to note – as we shall see in detail in the last part of the next chapter – that Duhem was bit of an exception and took metaphysics quite seriously. He in fact thought that if the physicist wants to understand the evolution of the subject and where it's headed, and have a purpose to her work, she must submit herself to metaphysics and affirm that physical theory is progressively latching on to an underlying reality. So he seems to have held that physics shouldn't interact with metaphysics when it comes to *practice*, but a contemplating physicist, setting aside immediate concerns of practice, will inevitably turn to metaphysics. As explained in section 6.3, I view this attitude of Duhem as rooted in a pragmatic rationality.

I now briefly look at the philosophies of Pearson, Poincaré, and Mach – as I said, in order to provide a contrast class – before delving into Duhem for a more detailed study in the next chapter.

5.2.2 Karl Pearson on Science, Knowledge, and Reality

Karl Pearson was a 19th century English mathematician, philosopher, an accomplished historian, and most importantly, the founding father of modern statistics. A contemporary of Pierre Duhem, Ernst Mach and Henri Poincaré, he was also an influential philosopher of science. A full blown antirealist about the existence of an external world, he believed that all of science, including its laws and formulas and observational claims, were constructs of the human mind. It is interesting thus, that he still had a working conception of truth in science – a truth that had nothing to do with scientific laws representing or reflecting any underlying reality, but that had entirely to do with "classifying facts and reasoning upon them". In this section I exposit his views on the characterization of science, the nature of facts and laws of science as expressed by him in his seminal work, *The Grammar of Science* (1892)

On the nature and role of Science

For Pearson, science was essentially about two things: classifying facts; and independence from the mind of the individual scientist. Science entails a dispassionate and proper classification, followed by identifying relations and sequences that follow a logical order. He says, “The classification of facts and the formation of absolute judgments upon the basis of this classification – judgements independent of the idiosyncrasies of the individual mind – essentially sum up the aim and method of modern science” (1892, 11) – and an individual who scrupulously carries out the above is said to have the “scientific frame of mind”. Pearson was greatly interested in the role of science in society and insisted that a scientific frame of mind was absolutely essential for good citizenship. According to him, such a frame of mind is not only found in scientists, but can and should be cultivated by anyone, by spending at least some time carefully studying one of the branches of science and understanding its method and approach. While the scientific frame of mind is necessary for good citizenship, Pearson is careful to note that not every accomplished scientist is automatically a good citizen until he carries the scientific method of classifying and organizing to all the fields he is engaged in.

Pearson goes on to make four important claims of modern science. I look at the first one in detail before going into the other three. Firstly, as above, education in modern science is absolutely essential for good citizenship, for “Minds trained to scientific methods are less likely to be led by mere appeal to the passions or by blind emotional excitement to sanction acts which in the end may lead to social disaster” (1892, 13). What sets science apart as far as relevance to society goes is *not* that it is more *useful* (than say philosophy or philology) in some tangible sense: science is a worthwhile endeavor because it is bound by reason and has a classification carried out by impersonal means. It employs and promotes values (such as being dispassionate while seeking knowledge) that are directly relevant to good citizenship. Pearson emphasizes that science is *method* rather than simply facts. The way science deals with facts is what makes science science: the examination, the classification, and coordination of every fact, past and present, with every other fact is the ultimate goal of science. The values of disinterestedness and meticulousness associated with such a method are what are essential to life at large.

So greatly does Pearson believe in the objectivity of the classificatory, and organizational power of science that he thinks that even if two individual scientists working entirely independently bring their work together, they will fit in with each other perfectly if they each indeed had a proper classification and were logically dealt with (this also shows that Pearson strongly believed in the uniformity of normal, rational human minds: this is a recurring theme in his writings and I shall come back to this at a later point). However, he thinks that the goal of science – which is "the complete interpretation of the universe" by means of a perfect and complete classification of all facts – is infinitely distant: it is an *ideal* goal. This, he says is because every new advance in science makes us observe new facts, which we failed to take into account earlier: this way, science keeps growing. Pearson's idea that science is essentially classification is very reminiscent of Duhem: Duhem also believed that the aim of physical theory was classification and devising "economical systems"; although Duhem and Pearson differed in what the science offers a classification of, and what it ultimately means. For Pearson, the classification is that of facts, and provided by laws, and this classification has nothing to do with realities in nature; whereas for Duhem laws classify phenomena, but more importantly, theory classifies laws, and this latter points to an ontological order in nature.

Further, Duhem too believed in an ideal, “perfect theory” that we could never get to, but approach asymptotically with each step of progress.

Pearson also goes quite deep into comparing science with other fields of thought and what distinguishes science from them. Mainly, the distinguishing point for him is proper classification, but he believes that any field can become scientific if it adopts the right kind of classification and organization into its system. For instance, he says of subjects that he thinks involve greater “bias of individual opinion” than traditional sciences like physics or chemistry:

> Our more thorough classification, however, of the facts of human development , our more accurate knowledge of the early history of human societies, of primitive customs, laws, and religions, our application of the principle of natural selection to man and his communities, are converting anthropology, folk-lore, sociology, and psychology into true sciences (1892, 19)

However, there are fields that simply are outside the domain of science: metaphysics is one of them. This, for him is because

> …the rules of methodical observation and the laws of logical thought do not apply to the facts, if any, which lie within such fields. These fields, if indeed such exist, must lie outside any intelligible definition which can be given of the word *knowledge.* If there are facts, and sequences to be observed among those facts, then we have all the requisites of scientific classification and knowledge. If there are no facts, or no sequences to be observed among them, then the possibility of *all* knowledge disappears” (1892, 18, emphasis as in original)

Thus, he makes the rather strong claim that only scientific knowledge is knowledge – one that conforms to the logical classification under discussion here. Pearson's complaint about the lack of proper classification in metaphysics is further explained by his following point: he argues that if indeed metaphysics were based on an accurate, scientific classification, then any normal, rational mind should understand it in the same way. However, metaphysicians disagree all the time; each metaphysician has his own set of beliefs. Thus, it is not possible that metaphysics follows a non-subjective classification, and is therefore outside the domain of science. For Pearson, the right attitude towards metaphysical questions is *agnosticism.* We must confess that we are ignorant when it comes to questions such as that of other consciousnesses – questions which science cannot address by means of its classification and logical sequencing. (As in the previous chapter, a pragmatic response to this might be that if we are to intelligibly engage in the activity of empathizing, we have to presume the existence of other consciousnesses. Affirming that other consciousnesses exist may or may not be good in the way of belief – if having correct beliefs is important, we should confess ignorance as Pearson says, but affirming it might just make pragmatic sense.)

Pearson proceeds to explain that there are two ways in which a field can be unscientific: either it can just have an improper classification, while its facts are legitimate in that they are about the observable world; or its very facts can be “unreal”. For instance, alchemy, whose subject matter was scientific enough to begin with, became chemistry – a science – with time, on embracing the scientific way of classification. On the other hand, there is no way that witchcraft can become science, for its very facts are “unreal creations of untrained minds” – they are absurd and unscientific. So as I understand, in the latter case, it doesn't matter whether or not a proper classification is possible: even if it is, the subject cannot become scientific since its “facts” can never be real. Now what about metaphysics? Pearson tells us that it does not have a proper classification, but are its facts “real” (but un-classifiable)? Or are its facts unreal to begin with? The question is important because if it is the former case, then like alchemy, metaphysics can hope to become scientific some day; but if it is the latter case, then it can never attain the scientific status. Pearson does not answer. He begins his discussion of this with the words “the rules of methodical observation and the laws of logical thought do not apply to the facts, if any, which lie within such fields.” (1892, 18): this indicates that he seems to be skeptical about calling the claims of metaphysics as “facts” although it is not clear why, for he doesn't say much about metaphysics and its claims other than that they are not scientifically classified. I shall return to this in the next section.

Several questions arise from the foregoing discussions: What exactly is this classification that Pearson repeatedly insists on, as essential to science? How does it classify facts? What *are* facts in the first place? Words Pearson consistently associates with “classification” are “logical order”, “relations” and “sequences”. It is not very clear what he means by these. In the simplest and most common sense, classification is understood as a grouping of things by similarity, like in the case of biology, which classifies organisms into various phyla. In physics, we can talk of similar phenomena being grouped together: repeating occurrences of an event are studied together, and a law or formula uniting those individual repeating occurrences would be formulated. At another level, there is a second type of classification: simple refraction and say, the aurora phenomenon both fall under the domain of Optics, and can hence said to be “classified” together. In yet another sense, there is theoretical unification of two or more subdomains, like quantum electrodynamics, which brings together quantum mechanics and electrodynamics. It is possibly all of these that Pearson had in mind when he spoke of classification. I return to the question of what facts meant for Pearson in a bit.

Although a firm idealist in the most traditional sense – someone who thought the universe actually varied from person to person and was entirely a construct in the mind of its observer – Pearson discusses *truth* in science: a truth divorced from any ties to an underlying metaphysical world. For him, truth is the goal of science, and he uses the term in a deflationary way: “The hard and stony path of classifying facts and reasoning upon them is the only way to ascertain truth” (1892, 20). For Pearson, truth then was the end point of all the laborious classification that science entailed and that was it. But crucially, what does such an end point look like? What exactly is the state of a perfect and complete classification? Unlike Duhem's idea of a perfect, natural classification of laws as an ideal goal (which we will look at in the next section), Pearson does not provide many clear, detailed answers.

Following the long discussion on Pearson’s first claim of modern science above, here are the other three. The second is that science helps form moral judgement. Pearson presents Weismann's idea that moral values are not acquired by children from their parents during their lifetime, but rather are passed on hereditarily. Science thus informs us of how morals are passed on, and helps people lead more morally good lives if they want their future offsprings to lead morally good lives too. Thirdly, he says the direct influence of pure science over practical life has been enormous. Even though a scientific endeavor might start off with no direct intention of having practical applications, it soon will greatly impact human life: “What at the instant of its discovery appears to be only a sequence of purely theoretical interest, becomes the basis of discoveries which in the end profoundly modify the condition of human life” (1892, 30). Lastly, science helps us relate with our aesthetic and imaginative side. Contrary to what it might appear like, science is not pure reason, devoid of imagination: “Disciplined imagination has been at the bottom of all great scientific discoveries” (1892, 30). Pearson holds that formulation of a succinct, elegant law surely involves imagination. Imagination is that crucial step after cataloging, that helps produce that one formula or law that brings the cataloged facts together. Of course, he says, we must follow that up with a careful scrutiny and self-criticism to check the limitations of the law or formula.

On Reality

In a nutshell, for Pearson reality is entirely constructed in the mind of the observer. It is a construct of present as well as past stored sense impressions. What we term consciousness, is in fact largely due to the stock of stored impressions and the way in which these condition the motor nerves when the sensory nerves pass on sensory information to them. Immediate sense impression is “that spark that kindles thought and brings into play stored impressions” (1892, 43). As briefly mentioned above, Pearson firmly believed that the minds and ways of conception and perception are the same in any two normal, rational individuals. Much of his philosophy is built on this faith in an inter-subjective objectivity. He says, “Two normal perceptive faculties construct the same universe.” (1892, 45) He reasons that if this were not the case, then “the results of thinking in one mind would have no validity for a second mind. The universal validity of science depends upon the similarity of the perceptive and reasoning faculties in normal civilized men” (1892, 45)

On the connection between the physical and the psychical, Pearson says we must declare ignorance and remain agnostic, for such a connection – if it exists – does not fall under the domain of science, and so there is no knowledge to be had there. For Pearson, to say that our thought or understanding of, say, the motion of a body, and the motion of the body itself, are one and the same, is not legitimate. We do not have the epistemic right to make such claims. In all likelihood, sense impressions produce some activity in the brain and this activity is recognized by each individual for himself/ herself under the form of some thought or understanding. But we cannot be sure of the relation between this internal understanding and the external phenomenon. However, one can find a vague Mach-ian strand (as we shall see in the next section) when Pearson says that at some future point science might be able to bring the two together. But for now, we have no choice but to remain agnostic.

Pearson explains that although all the content of the mind is ultimately based on sense impressions, there are many conceptions in the mind that are far removed from immediate sense impression and are often the results of long chains of inferencing from present and past stored impressions and in the process, classifying and analyzing them. For Pearson, these conceptions are no less scientific. As long as the ultimate source of the conception is sense impression, a conception can legitimately be scientific.

At this point I'd like to go back to the question of what count as facts for Pearson. Although he doesn't answer this directly even in the chapter titled 'The Facts of Science', I think we can safely assume that facts are not just about observational phenomena. As above, a fact produced by a conception that is not entirely rooted in sense impressions would still be legitimate. Also in light of his view that sociology and even folklore can be scientific with proper classification – of *facts* – it seems that facts can also belong to these other domains and not just the traditional sciences like physics. One thing we can be sure of is that Pearson was a great believer in the uniformity of the workings of normal, rational minds. So it was probably the case that he relied on the inter-subjective agreement of people to term something as a fact: such an understanding also accommodates his view that claims that are surely imagined and not shared by many people (like in witchcraft) are not facts. Of course, if this is the case then what a fact is will keep changing across different epistemic communities, and over time. But it does not seem that Pearson would have a problem with this.

Pearson strongly denies the possibility of having any objective knowledge at all of a world external to our phenomenal experience. Rather, he says, *we construct our universe* out of our sense impressions. Any notion at all we have of some external world, is one that *we project* outside of ourselves, and call it “external” or “real”. He discusses an illuminating analogy of a telephone operator: assuming that the operator has absolutely no access to the world outside of the phone, his world is one of sounds. If he were able to record some of the past messages received, his world would be nothing but a construct of incoming calls and past messages. What lies at the end of the wire – on the other side, he simply would never know. According to Pearson, the distinction between what is inside and what is outside of us is “entirely a function of the amount of the immediate sense-impression present at that instant.” So is Pearson saying that there is no external world at all – an ontological claim; or only that we cannot know anything about it – an epistemological claim? It is hard to say, but he makes an interesting point in response to this question: he says there is only one thing we might be able to say about an external world: that it produces sense impressions. But he is quick to clarify that even with this claim, we must be extremely careful, for we are associating causality – a feature of our phenomenal world – to a world that is not accessible to us. Strictly speaking, this is not a legitimate move to make. In sum, at the least, we must remain forever agnostic about any ‘noumenal’ world. (As we saw in the Chapter 4, a pragmatic response to this would be to say that in order to intelligibly engage in the activity of observation, we have to presume that there is an external world out there.)

This partly answers my previous question about Pearson's stand on metaphysical “facts”: the issue of ‘things-in-themselves’ being a metaphysical one, and Pearson demanding us to be agnostic about it, point to the possibility that he thought metaphysical facts were unknowable. And so possibly for Poincaré, what sometimes passed off as "facts" in metaphysics were unreal, as well as improperly classified.

Why is classification so important in science? Quite unsurprisingly, similar to Mach and Duhem, Pearson considers it an "economy of thought":

> Science deals with the contents of the mind, the 'inside' world, and the aim of its processes of classification and inference is precisely that of instinctive or mechanical association, namely to enable the exertion, best calculated to preserve the race and give pleasure to the individual, to follow on the sense-impression with the least expenditure of time and intellectual energy. (1892, 61)

Given a neat system of classification, our mind is able to rapidly and efficiently process new sense-impressions, combine them with the existing neatly and logically classified facts, and produce scientific knowledge.

On Laws

Laws are, again, human constructs about the world: "If nature for man is conditioned by his perceptive and retentive faculties, then natural law is conditioned by them also." (1892, 73) Pearson goes so far as to say that a law is in fact a creation, an invention, rather than a discovery: it is the brainchild of one individual scientist who wrote it down as a *mental shorthand* for otherwise lengthy descriptions of the sequence of events in our sense-impressions. For Pearson, a fundamental law of nature is one that embraces as wide a range of phenomena as possible, and as simply stated as possible. It is worth emphasizing at the same time, that Pearson was as inter-subjectively-objectivist about facts and laws of science as he was anti-objectivist about the existence of a world external to us. i.e., although the world for him is entirely a product of human reason and imagination, facts and laws of science have inter-subjective objective validity among normal, rational, logical minds. Laws are considered to be universal in that they describe – conditioned by our "perceptive faculty" – the world of phenomena, which must be the same for everyone. Laws are also conditioned by our "reflective faculty" which enables the processes of inference and logical classification, which again must be the same for all persons.

Pearson argues that laws don't *explain,* but only describe:

> The law of gravitation is a brief description of *how* every particle of matter in the universe is altering its motion with reference to every other particle. It does not tell us *why* particles thus move; it does not tell us *why* the earth describes a certain curve around the sun. It simply resumes, in a few brief words, the relationships observed between a vast range of phenomena. It economizes thought by stating in conceptual shorthand that routine of our perceptions which forms for us the universe of gravitating matter. (1892, 87, emphasis as in original)

Just like Duhem who held that explanations were the job of metaphysics and not of physics, Pearson says, laws never explain "the routine of our perceptions, the sense-impressions we project into an 'outside world'." (87) (Of course, Duhem and Pearson varied in their attitude towards metaphysics: as we shall see, the former took metaphysics way more seriously than the latter). Laws thus don't "rule" nature, but only describe phenomena; and the replacement of one law by another is part of the natural course of scientific progress.

An alternative to professing ignorance in domains outside of science – like metaphysics – is often suggested: it is easy to assert that where science produces no positive knowledge, we hypothesize, and hold on to it until science proves otherwise. But for Pearson, this involves "undisciplined imagination". Such hypotheses, for Pearson, constitute no knowledge at all. "Driven from one stronghold of ignorance, those who delight in the undisciplined imagination rather than in positive knowledge, only seek refuge in another." (1892, 97) Pearson distinguishes two kinds of ignorances: one about phenomena, that falls within the domain of science, and another one, in an unknown, unknowable world that we project outside of ourselves. It is only the former that we can work to eliminate, in due course of scientific progress. The latter, we must dutifully ignore. Clearly, Pearson was a thoroughgoing antirealist, with no room for any minimal realist interpretation unlike Mach or Duhem. His antirealism seems to have been nothing short of traditional idealism.

5.2.3 Henri Poincaré on Truth in Physics

Henri Poincaré, a 19th century mathematician, physicist, and philosopher, frequently talked about the "true relations between real objects" reflected by equations in physical theories. At the same time, he seems to have been an antirealist in the metaphysical sense in that he denied the possibility of access to a world completely external to the human mind. I explore these ideas here, based on Poincaré's *Science and Hypothesis* (1952) and *The Value of Science* (1958).

Poincaré on Experiment, Mathematical Physics, and Truth

At the beginning of the chapter ‘Hypotheses in Physics’ in *Science and Hypothesis* (1952), Poincaré says that experiment is the sole source of truth since it is the only source of certain knowledge: “it alone can teach us something new; it alone can give us certainty” (1952, 140). He then asks, if only experiment can lead to truth, what of mathematical physics? What is its role? Its role, he says, is to generalize the results of observation. Mathematical physics is what provides *method* to science, which would otherwise consist of just plain observational facts like a “heap of stones”. Science is more than an accumulation of isolated facts: ordering and organizing are essential to its character, and this is precisely the role of mathematical physics. Science, Poincaré says, is like a library, with experimental physics purchasing and providing the resources. What mathematical physics then does, is draw up a catalogue for them and make generalizations:

> If the catalogue is well done the library is none the richer for it; but the reader will be enabled to utilise its riches; and also by showing the librarian the gaps in his collection, it will help him make a judicious use of his funds, which is all the more important, inasmuch as those funds are entirely inadequate. That is the role of mathematical physics. (1952, 144)

An interesting inference follows: if truth only belongs to the experimental domain, and if mathematical physics is still essential to science, then it seems that for Poincaré, science is more than truth: the method that mathematical physics provides is what protects the integrity of science. What relation does mathematical physics bear with truth though? Although Poincaré says experiment is the sole source of truth, it turns out that he things mathematical physics also has an important relation to truth. I return to this later.

For Poincaré, an experiment can never be repeated, strictly speaking: an observed fact is never exactly repeated. But the facts produced by various repeated instances of the experiment must be joined together by mathematical physics if we are to be able to generalize and predict. If the results of different experimental runs are represented by points, it is mathematical physics that joins them by means of interpolation: it draws a line that does not pass exactly through them, but that passes near and between them. Detached facts do not satisfy us, we thus look for connections and generalizations so that we are able to make predictions. But precisely since a prediction is often a result of this kind of an interpolation, we can never be absolutely sure of its certainty: experiment might very well falsify it. But, "…the probability of its accuracy is often so great that practically we may be content with it. It is far better to predict without certainty, than never to have predicted at all." (1952, 144)

For Poincaré, all generalizations are hypotheses that must be verified by experiment. But if a hypothesis is disproved by experiment, there is no reason to despair. On the other hand, it is a desirable outcome for it makes the experiment a decisive one.

Further, "if this experiment had been made by chance without the hypothesis, no conclusion could have been drawn; nothing extraordinary would have been seen; and only one fact the more would have been cataloged, without deducing from it the remotest consequence." (1952, 151)

Poincaré interestingly invokes a very Duhemian holism. He says that while using mathematical physics to formulate and organize hypotheses, we shouldn't unnecessarily "multiply hypotheses indefinitely": "If we construct a theory based upon multiple hypotheses, and if experiment condemns it, which of the premisses must be changed? It is impossible to tell. Conversely, if the experiment succeeds, must we suppose that it has verified all these hypotheses at once? Can several unknowns be determined from a single equation?" (1952, 152) Therefore, he says, it is safer to play with fewer hypotheses.

What is the relation between mathematical physics and truth? For Poincaré, the *relations* occurring in a (empirically successful) theory are indicative of true relations between things in reality. The actual nature of the underlying reality described by physics is inaccessible to us. As he says in the preface of *Science and Hypothesis*, when we are doing physics, our concern is only with relations rather than the things themselves. What does he mean by relations? We get an idea of this from the following passage:

> That a given periodic phenomenon (an electric oscillation, for instance) is really due to the vibration of a given atom, which, behaving like a pendulum, is really displaced in this manner or that, all this is neither certain nor essential. But that there is between the electric oscillation, the movement of the pendulum, and all periodic phenomena an intimate

> relationship which corresponds to a profound reality; that this relationship, this similarity, or rather this parallelism, is contained in the details; that it is a consequence of more general principles... (1952, 161)

So for Poincaré, apparently disparate phenomena like electric oscillation and pendulum motion are brought together by the general principles of mathematical physics. We don't and can't know anything about the true nature of electric oscillations, but we know that it is related to pendulum motion as the equations of the two are structurally similar. This idea of structural similarity of equations across theories indicating real relations can be gleaned from the following. Talking of theories of dispersion, Poincaré says,

> But the remarkable thing is, that all the scientists who followed Helmholtz obtain the same equations, although their starting-points were to all appearance widely separated. I venture to say that these theories are all simultaneously true; not merely because they express a true relation— that between absorption and abnormal dispersion. In the premisses of these theories the part that is true is the part common to all: it is the affirmation of this or that relation between certain things, which some call by one name and some by another. (1952, 162)

As mentioned earlier, Poincaré sees the role of mathematical physics as organizing a heap of stones, or cataloging of books. So there is order and organization in science, and this reflects an underlying order and organization in the form of relations between things in nature, and this is all we can know of nature.

Further, according to Poincaré, this order is preserved across theories:

> This Fresnel's theory enables us to do today as well as it did before Maxwell's time. The differential equations are always true, they may be

> always integrated by the same methods, and the results of this integration still preserve their value. It cannot be said that this is reducing physical theories to simple practical recipes; these equations express relations, and if the equations remain true, it is because the relations preserve their reality. (1952, 161)

According to Poincaré, theories are mere "images" reflecting objects in nature and it doesn't matter what image we impose upon nature as long as the *relations* captured in the images are true: "The true relations between these real objects are the only reality we can attain and the sole condition is that the same relations shall exist between these objects as between the images we are forced to put in their place." (1952, 161) These views of Poincaré formed the precursor to modern structural realism attributed to John Worrall. Once again this is all very similar to Duhem's views on physical theory – Duhem also thought that theory gets right only the relations among things. I look at this in detail in the next chapter.

As we see here, Poincaré has invoked a second conception of truth: while the first one solely had to do with experiment, here we see that it has to do with theory – which constitutes mathematical physics – although truth here is only one at the level of relations. The two ideas on truth seem incompatible and it seems strange that given Poincaré's leanings towards the truth of relations among things reflected in mathematical physics, he would claim that experiment is the sole source of truth.

Although it might seem – going by the above discussion on relations in theory corresponding to real relations – that Poincaré was a realist, perhaps a structural one at that, the following passage in *The Value of Science* (1958) show clearly that he was a antirealist with respect to the existence of a mind-independent world:

> Does the harmony the human intelligence thinks it discovers in nature exist outside of this intelligence? No, beyond all doubt, a reality completely independent of the mind which conceives it, sees or feels it, is an impossibility. A world as exterior as that, even if it existed, would for us be forever inaccessible. But what we call objective reality is, in the last analysis, what is common to many thinking beings, and could be common to all; this common part, we shall see, can only be the harmony expressed by mathematical laws. It is this harmony then which is the sole objective reality, the only truth we can attain..." (1958, 14)

(As we saw in the last chapter, we could take a pragmatic approach and not deny the existence of an external world in order to intelligibly engage in the activity of observation.)

If a mind-independent reality is impossible, and even if it were possible, inaccessible to us, how to make sense of Poincaré's view that relations expressed by equations in physics correspond to a "profound reality"? The answer lies in the last part of the passage quoted above: he believes in the inter-subjective objectivity of human minds. So the profound reality that corresponds to the theoretic relation between electric oscillation and pendulum motion must be an inter-subjective reality, probably the phenomenal world as we understand and talk about it.

How to reconcile Poincaré's first and second ideas on truth? Poincaré first claimed that experiment is the sole source of truth – if he means here truth of underlying reality, we've seen that it would be inter-subjective truth. But it is not clear why he says experiment is the sole source of (inter-subjective) truth, when later he goes on to say that the relations that mathematical physics gives us correspond to an (inter-subjective) truth about underlying relations between things in nature. The former claim about experiment being the sole source of truth seems to be an aberration and doesn't line up with the bulk of Poincaré's views on physics and truth. Setting that aside then, Poincaré can be thought of as a realist – committed to truth in an inter-subjective sense – with respect to the relations between realities that mathematical physics proposes, but a metaphysical antirealist, for while he thought physical theory truly represents the relations in the underlying ontology of nature, he held that we cannot have access to any mind-independent reality. The reality (of relations between things) that mathematical physics reveals is not a mind-independent reality, but only a reality shared by "many thinking beings" (1958, 14).

5.2.4 On Ernst Mach's Economy and (Anti) Realism in Science

Ernst Mach was a very influential figure in physics and philosophy of science in the 19th century. The focus of this section will be his philosophical ideas on the role, status, scope, and goals of laws and theories in science; and his (anti)realist worldview. Specifically, I review his views on the economy of the organizational structure that is science, and draw inferences about how (anti)realist they were about this. I ask: Did he believe that this economical organization represented any underlying metaphysics of nature at all? Did he believe in the existence of a world outside our thought?, etc. I will follow E.C. Banks in arguing that we can indeed have a realist interpretation – albeit quite minimal and unconventional – of Mach's views on Science.

Ernst Mach on the Economy of Thought and Physical Inquiry

Ernst Mach discusses at great length, how we as humans economize thought, and in general any body of knowledge that we construct. In the 1882 Vienna Imperial Academy of Sciences anniversary meeting address, 'The Economical Nature of Physical Inquiry' he starts by drawing an analogy with how our mind routinely completes a partial image: for instance, how on seeing a rind, we immediately associate it with a fruit although the fruit may not be seen. Mach holds that such a psychological act is the result of "the economy of our organism rooted not less firmly than motion or digestion". (1898, 190). This, is the first step in Machian economy-talk. As I see it, there are three important facets to Mach's theory of economy in science. First, he argues that this sense of economy is something (biologically) *a priori*: it is not entirely experiential – a large part of it is "instinctive". Further, as above, he considers it to be as basic to humans as motion or digestion. He says,

> Our instinctive knowledge, as we shall briefly call it, by the virtue of the conviction that we have consciously and intentionally contributed nothing to its formation, confronts us with an authority and logical power which consciously acquired knowledge even from familiar sources and of easily tested fallibility can never possess. (1898*,* 190)

Next, he makes the point that this instinctive faculty of making economical associations and constructions is cumulatively passed on to each generation from its predecessors – it is in this sense that economizing is a priori. It is not rational a priori that he means, but a biological a priori: for instance a particular principle or idea of economizing may be a priori to one generation since its predecessors have already come up with it, but might have been a posteriori to the predecessors themselves. It is very important for Mach that we acknowledge that a lot of "economizing labor" is saved due to the extant work in the same direction by previous generations. Finally he says that the "communication of experience" is economized: we capitalize on the descriptions and vocabularies that already existed in order to convey a scientific idea (natural language itself being a case in point). Further – and here comes the crux – he says we aim for concise and economical descriptions in our scientific communications. "to save the labor of instruction and of acquisition" (1898*,* 193) To him, s*cientific laws embody this economy* – for they generalize over, and unify, several different phenomena – and they also help economize our thought, for they aid in efficient understanding of phenomena. He makes the point that we make sense of a phenomenon, say the ascent or descent of a body under the action of gravity, with the backdrop of the laws of motion: we don't comprehend it as an isolated instance that needs a separate explanation – which would be a "burdening of the memory" – but rather as an instantiation of the same laws we already know; laws which economize our scientific communication and thought. Laws and formulae in science group together similar phenomena and help predict the outcomes of new instances of those/ related phenomena[13]. To him, mathematics is the epitome of

[13] It is far from clear what Mach thought was the exact relation between laws and the

economy: by its neatly organized system of signs and formulae for easy calculations, it helps save all “unnecessary thought” and “mental operations”(1898, 195).

The second facet outlined above – the passing on of economy from generation to generation – is worth elaborating for it is important to Mach. Mach sees economy in science as a means to intellectual self-preservation. As above, he stresses that the labor of science, and thought in general, of one individual is greatly reduced by the work done by her predecessors. While it is never impossible to solve, say, an arithmetical problem by direct counting; it can be solved way more quickly and efficiently if we rely on the operations already devised and put in place to its end. He makes the following analogy:

> Just as a single human being, restricted wholly to the fruits of his own labor, could never amass a fortune, but on the contrary the accumulation of the labor of many men in the hands of one is the foundation of wealth and power, so, also, no knowledge worthy of the name can be gathered up in a single human mind limited to the span of a human life and gifted only with finite powers, except by the most exquisite economy of thought and by the careful amassment of the economically ordered experience of thousands of co-workers. (1898,198)

Mach had a biological conception of economy: he almost thought of economy of thought as hereditary. In the inaugural address on assuming the Rectorate of the University of Prague in October 1883, he develops a labored and long drawn analogy between Darwinian evolution and mental adaptation. Although wary of carrying the analogy too far and too literally, Mach insists that the “general imprint of evolution and

---

phenomena they represent, although he makes one thing clear: that laws can never capture the phenomena they purport to represent, in their entirety. I will pass over this issue for the moment and return to it in my discussion of his realism in the next section.

transformation must be noticeable in ideas also".(1898, 218) He clarifies this analogy with the following example: just like if a vertebrate has to swim or fly it doesn't develop an entirely new pair of wings or fins but rather *adapts* to the new environment with minimal changes; so too do humans, with their ideas. If we encounter a new phenomenon, our first instinct is to fit it within the framework of our existing ideas with minimal modifications. We constantly build on existing ideas and without the treasure of past knowledge built over generations, we would be unable to economize thought and would be intellectually quite unevolved. Science is a continuous tightrope walk between the judgements amassed over time, and new judgments pronounced on new, previously unobserved phenomena. Again, in a German Association of Naturalists and Physicians address at Vienna in 1894, Mach explains how ideas are arrived at by *comparison* to older ones. For instance as he says, "The concepts of force, mass, and work are then carried over, with appropriate modifications, to the phenomena of electricity and magnetism".(1898, 249) What seems to be unique about Mach – and what, as E.C Banks says annoyed his phenomenologist critics the most – is that "he considered the disposition toward economy and law to be a *pre-rational* need" – something biologically ingrained – "in advance of normative frameworks of thought and language, rather than as a priori rational ordering". (2004, 34)

While claims based purely on simple observation, like claims about the color or shape of on object, Mach calls *direct descriptions*, claims more theoretically embedded are *indirect descriptions*. Subsuming more direct descriptions within indirect descriptions helps with economizing in science. A theoretical idea, like the wave nature of light, thus, had for Mach, the power to "extend the fact, and enrich it with features which we are first induced to seek from such suggestions, and which are often actually found. It is this rapidity in extending knowledge that gives to theory a preference over simple observation" (1898, 241). This preference was however, entirely quantitative, and not qualitative for Mach. I return to this point in greater detail in the next section.

Having briefly surveyed Mach's views on economy in science, I take them up in more detail in the next section which is about his (anti) realism. What kind of a scientific realist or antirealist was Mach, and what can we know about this from his elaborate theory of economy? This is the question I address below.

Ernst Mach's (Anti) Realism

Throughout his discussion of economy of thought; Mach makes it clear that none of the economizing that the mind acquires from predecessors and carries out by itself, and none of the economical organization afforded by physics has any correspondence with how things are in nature itself. To begin with, an important aspect of Mach's philosophy was that events aren't actually regular in nature although we might represent them by regularity laws: "Nature exists only once." (1898, 199) For Mach, it is only our limited mental faculty that yearns for patterns and regularities, that produces like, recurring events; not nature. Laws are thus inevitably inexact and are only economical constructions: "In reality the law always contains less than the fact itself, because it does not reproduce the fact as a whole but only in that aspect of it which is important for us, the rest being either intentionally or from necessity omitted." (1898, 193) Thus, he says that absolute forecasts based on laws are not very significant in science[14]. In fact, he quite explicitly states that the theory and formulae we use to talk about phenomena have nothing to do with the phenomena themselves:

> Although we represent vibrations by the harmonic formula, the phenomena of cooling by exponentials, falls by squares of times, etc., no one will fancy that vibrations in *themselves* have anything to do with the circular functions, or the motion of falling bodies with squares. It has simply been observed that the relations between the quantities investigated were similar to certain relations obtaining between familiar mathematical functions, and these *more familiar* ideas are employed as an easy means of supplementing experience. (1919, 492; emphasis as in original)

So for Mach, physical theory was merely a means of economizing thought and ideas, and a mental construct.

This is similar to Pierre Duhem's views on physical theory according to which theories stored laws in "condensed representations" and "economical systems" (1906/1954, 23) (and laws were in turn symbolic representations of phenomena). One difference between Mach and Duhem though – as Thomas Hickey (1995) points out – is that while the former only stressed on the economical aspect of theories, the latter had an almost equally elaborate argument about the representational role of theories: theories are not identical with empirical laws, but represent them by means of a symbolic structure. An even more significant difference is that Duhem believed in a metaphysics of physical

[14] Among contemporary views, this is reminiscent of Nancy Cartwright's (1999) arguments that laws are simply not accurate as they can never capture the complexity of individual phenomena – the world is "dappled".

theory, quite separate from physics. His belief was that as long as one is doing physics, one must not question the truth of physical theory, but that once one steps out of physics and thinks about all the progress of physical theory – its success at predicting and anticipating experimental results – one is forced to think that physical theory is at the least converging onto some metaphysical truth about nature. Mach on the other hand shied away from making any connection at all, between physical theory and truth of any kind.

It thus seems that Mach was by no means a scientific realist as the term is understood today: he did not take scientific theories to have anything to with the phenomena they represent. But did he believe in the existence of a world external to our thought? If so, what was its nature? This is a much tougher question. Mach does not answer this question in any direct way, but there are many instances in his writings that indicate that he was not a realist about the existence of an external world either. A discussion of Mach's antirealism would not be complete without his anti-atomistic views. He says in a response to Max Planck in 'The Guiding Principles of My Scientific Theory of Knowledge and Its Reception by Contemporaries', "If belief in the reality of atoms is so important to you, I cut myself off from the physicist's mode of thinking. I do not wish to be a true physicist, I renounce all scientific respect – in short; I decline with thanks the communion of the faithful. I prefer freedom of thought." (quoted in Pojman, *The Stanford Encyclopedia of Philosophy* 2011) While Mach was ready to use the atomic theory of his day as long as it was fruitful, he was opposed to committing to any metaphysical reality of atoms.

Mach's anti-atomism was a reflection of his larger, apparently antirealist perspective. Consider his explanation of our conception and understanding of identity: although an object may change in many aspects, our mind retains certain properties it takes to be constant so that it can repeatedly identify the object as being the same in a wide range of situations. He then goes on to say that we seem to have formed a notion of the object independent of its attributes; of a "thing-in-itself", while our sensations are taken as mere "symbols" of the attributes of this thing. Here come Mach's crucial words on the realism of these things: "But it would be much better to say that bodies or things are compendious mental symbols for groups of sensations – symbols that do not exist outside of thought." (1898, 201)

According to Mach, when we refer to the white metal, the liquid it becomes on being heated, and the violet vapors it turns into on being further heated, all by one name, sodium, what we are essentially doing is economizing thought by conveniently grouping them together as different states of the same substance. Mach says, "But more than a compendious economical symbol for these phenomena, that name and thought is not." (1898, 202) Further, Mach says soon after discussing this example, that a body is a "compound of light and touch sensations" (1898, 203) indicating that it is but mental. As we shall see below though, Mach uses the word "sensations" only provisionally.

Mach's conception of the external world is that it is made of what he calls "*elements*". And what are these elements? Mach says: "Properly speaking the world is not composed of "things" as its elements, but of colors, tones, pressures, spaces, times, in short what we ordinarily call individual sensations."(1919, 483) But Mach was not entirely happy with calling elements sensations because in the term, "an arbitrary, one-sided theory is embodied" (1898, 209). This is an important point that further explicates Mach's (anti) realism: what he means by "one-sided" is the human, mental side. While "sensations" refer to the mental, "elements" is a more general term and can refer to things non-mental as well. This is important because it goes to show that Mach did believe in the existence of *something* outside of our thought. What really are elements then?

Elements are the same as sensations, *when* they are taken in the context of *human experience*. i.e., when we speak of a body being perceived by us, we perceive the component elements of the body as sensations. So in this sense, he says, *to us*, the world *is* our sensations. But with the subject removed from the picture, he thought the world itself is composed of elements. What these elements are, independent of the human mind, he unfortunately does not say much about. Ultimately Mach's firm belief is that a deep investigation into the psychology and physiology of these sensations will reveal much about the "true, real elements of the world". (1898, 212) He says,

> But we certainly shall wonder how colors and tones which were such innermost parts of us could suddenly get lost in our physical world of atoms; how we could be suddenly surprised that something which outside us simply clicked and beat, in our heads should make light and music; and how we could ask whether matter can feel, that is to say, whether a mental symbol for a group of symbols can feel? (1898, 213)

Mach's dream was to devise one whole, complete theory encompassing both physics and physiology; one that brings together the physical and the psycho-physiological.

What is ambiguous, yet interesting about Mach's metaphysical realism is that although he refers to “body” and “matter” as mere symbols, he does seem to think that there is something in the world outside of the mental realm. He explicitly states that we need not be bound by the idea that elements are nothing more than our sensations. (1898, 209) Further, as in the above quote, he believes that there are certain “true, real elements of the world” which a thorough probe into physiology will some day reveal. Given these views, what sort of an ontology can be afforded to Mach if we want to consider him a metaphysical realist of some sort? Certainly, it would be an ontology built on *elements*, but it is of course not very clear what such an ontology would look like.

E.C. Banks (2004) puts forward an interesting realist interpretation of Mach based on his “elements”. His view is that Mach's position on economy and his antirealism about laws was rooted in his *realism* about *particulars.* i.e. Mach believed that while laws could not capture the diversity of individual, particular events – while the regularity in the laws was not real – the particular events, and the things carrying them out, themselves were. But what was the nature of these particulars then? Banks produces a quite revealing quote from Mach from an unpublished lecture at Prague:

> Sensation is a general property of matter, more general than motion. Let us seek to set down this proposition clearly. An organism is a set of molecules. Electrical currents run into the interior and come back again

> into the muscles. Everything is physically explainable. But not that the person should have sensations. What we can investigate physically is always merely physical. We feel no sensation. And yet the human being senses. The material flows forward through and through him. The old departs; the new comes in. We have therefore the problem of finding something fundamentally new in the whole that is not in the parts. We escape this difficulty when we consider sensation as a general property of matter. (quoted in Banks, 2004, 43)

So while earlier we saw that it seemed like Mach wanted to distinguish elements from sensations and held that matter is constituted by elements, here it looks like he took sensations to constitute matter. Banks gives another quote from Mach's 1863 *Lectures on Psychophysics* where Mach wrote that "material atoms could be constructed out of physical qualia similar to sensations" (Banks, 2004, 43) Mach is further said to have thought that sensations could be attributed to matter by analogy: humans and animals having sensations could be extended to inanimate matter as well. But Mach was happy to stop with the idea that elements and sensations had a mutual dependence and not insist unequivocally that matter was comprised of sensations, until future experiments so revealed.

Hence, Mach can be thought of as a realist, but in a very minimal, unconventional way: he believed in the existence of a world that consisted of matter composed of sensations, which Banks characterizes as some "tiny neutral qualities". While theories and laws only helped with economy of thought and did not say much about nature, some day, in a future physio-psycho-physics, Mach hoped for a holistic reconciliation of nature and our experience and perception of it.

5.3 Conclusion

All of Pearson, Poincaré, and Mach were antirealist about science – they all thought theories serve to economize thought and do not get to any underlying, metaphysical reality. They also were metaphysical antirealists in that they did not think an external, mind-independent world exists. They were all, in general, anti-metaphysics. Pearson thought we construct our universe out of our sense impressions; Poincaré thought theories get right the relations between things in a shared, inter-subjective reality (presumably the phenomenal world as we understand and talk about it) and that a mind-independent reality is an impossibility; and Mach thought all that exists are sensations. Duhem, as we'll see was an exception to these 19th century scientist-philosophers. It doesn't seem that he was a metaphysical antirealist, and he thought that the physicist justifiably affirms that the classification of laws that physical theory afford progressively mirrors a natural classification.

References

- Banks, E.C (2004) The Philosophical Roots of Ernst Mach's Economy of Thought *Synthese* 139 23-53
- Cartwright, Nancy (1999) *The Dappled World* Cambridge University Press
- Rollet, Laurent (2014) 'Portrait of Henri Poincaré as a Young Philosopher: The Formative Years' in de Paz, Maria and DiSalle, Robert eds. *Poincaré, Philosopher of Science*, The Western Ontario Series in Philosophy of Science (79) Springer 3-24

- Duhem, Pierre (1906) *The Aim and Structure of Physical Theory.* (Repr. 1954) (Philip P. Wiener, Trans.) Princeton, NJ: Princeton University Press

- Duhem, Pierre (1996) *Essays in the History and Philosophy of Science* Hackett Publishing Company

- Hickey, Thomas J (1995) *History of Twentieth Century Philosophy of Science* Thomas J. Hickey

- Kloppenberg, James (1986) *Uncertain Victory: Social Democracy and Progressivism in European and American Thought, 1870-1920* Oxford University Press

- Mach, Ernst (1898) *Popular Scientific Lectures* The Open Court Publishing Co.

- Mach, Ernst (1919) *The Science of Mechanics* The Open Court Publishing Co.

- Pearson, Karl (1892) *The Grammar of Science* J.M. Dent and Sons Ltd

- Poincaré, Henri (1952) *Science and Hypothesis* Dover Publications, Inc.

- Poincaré, Henri (1958) *The Value of Science* Dover Publications, Inc.

- Pojman, Paul (2011) 'Ernst Mach' *The Stanford Encyclopedia of Philosophy* Winter Ed. Edward N. Zalta (ed.) http://plato.stanford.edu/archives/win2011/entries/ernst-mach/

- Stanford, Kyle (2006) 'Instrumentalism' in Sarkar, Sahotra and Pfeifer, Jessica eds. *The Philosophy of Science: An Encyclopedia* Routledge, 400-405

6. Duhem's Philosophy of Physical Theory

6.1 Introduction

In this chapter I take up a detailed study of Duhem – in particular, his idea of natural classification. In Section 6.2 I ask if Duhem was a structural realist – whether he was concerned with structure as is understood in philosophy of science today, post John Worrall's (re)introduction of the view in 1989. Duhem famously held that physical theory classifies experimental laws, and that we are compelled to believe that this classification progressively tends to a *natural* classification: one that reflects a real metaphysical order. Many scholars have interpreted Duhem as a structural realist. I contend that the bulk of Duhem's views don't line up with structural realism if structure, as John Worrall takes it to be, is the mathematical form of equations. Duhem's "structure" – if we can call it so – rather meant the classification of experimental laws brought about by the equations of theory.

Then in Section 6.3 I look into the rationale behind Duhem's natural classification and the interpretation of Duhem's attitude towards physical theory. Based on his view that the classification of experimental laws yielded by theory progressively approaches a natural classification – a classification reflecting that of underlying realities – Duhem has been construed as a realist of sorts in recent literature. Here I argue that Duhem's idea of natural classification was an intuitive idea in the mind of the physicist that had to be affirmed in order to justify the physicist's pursuit of theory. I situate Duhem in Chang's scheme of epistemic activities and ontological principles discussed in Chapter 4, and argue that Duhem's positive attitude towards the theoretic classification of laws had rather to do with the pragmatic rationality of the physicist. Affirming that the theoretic classification of laws was approaching a natural classification rationalized the practice of physical theory for the physicist and in this way made her practice more *understandable* to her, regardless of how the practice 'actually' or 'objectively' works. In a way then, I extend the Changian ideas on understanding *in* science from Chapter 4 to philosophical understanding – specifically, understanding *of* science by the (physicist and) philosopher of science.

6.2 Duhem, Natural Classification, and Structure

6.2.1. Duhem on physical theory: an overview

According to Duhem, physical theories *represent*, *organize and classify* experimental laws. Many thinkers go further, maintaining that theories explain, in a strong sense: to describe real, underlying causes of sensible phenomena, or as Duhem says, "strip reality of the appearances covering it like a veil, in order to see the bare reality itself" (1906/1954, 7). Duhem rejects this. Such explanation, for him, is the business of *metaphysics.* The business of physics is only to study phenomena, find experimental laws, and organize and classify these laws. But why can't physics study the causes? Because it is beyond its means: the experimental method of physics does not have the resources to provide any positive metaphysical knowledge.

For Duhem, physical theories provide an "economy of thought" and serve to "store" the otherwise intractably enormous number of experimental laws in "condensed representations" (1906/1954, 23). "[I]t matters little" for him whether the operations performed to combine various hypotheses together "do or do not correspond to real or conceivable physical transformations". (1906/1954, 20) All that matters is that theories are consistent with the laws they represent.

And importantly, as above, Duhem held that theories *classify* experimental laws. Here is a particularly revealing passage in this regard, call it CL. Duhem says of theory,

> **CL**: ... alongside the laws which govern the spectrum formed by a prism it arranges the laws governing the colors of the rainbow; but the laws according to which the colors of Newton's rings are ordered go elsewhere to join the laws of fringes discovered by Young and Fresnel; still in another category, the elegant coloration analyzed by Grimaldi is considered related to the diffraction spectra produced by Fraunhofer. The laws of all these phenomena, whose striking colors lead to their confusion in the eyes of the simple observer are, thanks to the efforts of the theorist, classified and ordered. (1906/1954, 24)

Is this classification indicative of features of metaphysical reality? Not in any straightforward sense, for Duhem. This is particularly clear in his 'Physics and Metaphysics' in *Essay in the History and Philosophy of Science* (1996): "Laws of physics retain exactly the same sense when a theory connects them as when they are dispersed and isolated." (1996, 36) Theorizing – i.e. representing and classifying laws economically – has no effect on the character of physical science: "It remains physics; it does not become metaphysics." (1996, 36) He continues, "A classification .... is not a judgement. It can be convenient or inconvenient, good or bad; it cannot be true or false." (1996, 37)

Despite this seemingly instrumentalist view that is often attributed to him, Duhem undoubtedly had a positive attitude[15] towards the classifying power of theory. Not only did he find it "beautiful", but greatly valued it for its elegance and perfection of economical representation. He affirms that we are persuaded to believe that the classification tends to reflect a *natural, underlying order*; that the relations among experimental laws established by theory "correspond to real affinities among the things themselves" (1906/1954, 26). One argument for this was that if the classification brought about by physical theory were a purely "artificial" system having nothing to do with a real underlying order, then it would be a "marvelous feat of chance" (1906/1954, 28) that it anticipates new laws mostly correctly, and so frequently gets right novel predictions. But if on the other hand we took it to increasingly reflect a metaphysical order, then it would be no surprise or accident if it succeeded in the ways just described. For Duhem, theory has as its "limiting form" a *natural classification* (1906/1954, 293): a perfect, ideal classification of experimental laws. Such a classification is *natural* in that it

[15] Duhem had a lot to say about the basis for, and the epistemic scope and limitations of, such an attitude. Some scholars have seen this attitude as grounded in a kind of a no-miracles argument. Here I don't go into any questions about Duhem's rationale behind this attitude – I look at that in the next section. I'm only interested in the object of this attitude.

perfectly mirrors an ontological order: "...the more complete it becomes, the more we apprehend that the logical order in which theory orders experimental laws is the reflection of an ontological order" (1906/1954, 26). This idea recurs in several passages in Duhem (1906/1954), and this will be a central point for the remainder of this section: we are compelled to believe that theoretic classification of experimental laws approaches a natural classification that mirrors an ontological order.

6.2.2. Was Duhem concerned with structure?

As above, Duhem says that we apprehend the (possible) reality of some relations between the laws governing phenomena established by the classification, in that we feel the relations among the laws established by theory tend to reflect a real metaphysical order: "...physical theory through its successive advances tends to arrange experimental laws in an order more and more analogous to the transcendent order according to which the realities are classified..." (1906/1954, 297) But he doesn't think we can apprehend, in a similar way, anything about the nature of the realities behind the phenomena in the metaphysical realm: "Without claiming to explain the reality hiding under the phenomena whose laws we group, we feel that the groupings established by our theory correspond to the real *affinities* among the things themselves." (1906/1954, 26, emphasis mine) It is clearly this view – this positive attitude towards taking the relations between experimental laws captured by theory to approach ontological affinities or groupings – that has given rise to structural realist interpretations of Duhem. Favoring the idea that theory progressively gets right the 'affinities' of things at the metaphysical realm and not the nature of the things (or anything else) does after all prima facie sound similar to being realist about structure but not about what is structured. I want to alert the reader though that in this section, I am not concerned with whether Duhem was a structural (or any other kind of) *realist* (I look at that in the Section 6.3), but only with whether the object of his positive attitude towards physical theory – in thinking that it was approaching an ontological order – was structure. So without getting into questions of realism, I shall deal with whether Duhem's arguments about natural classification warrant a structuralist interpretation.

It will serve us well to first recall the central ideas of modern structural realism. John Worrall (1989), who first developed and defended modern structural realism, follows Henri Poincaré and asserts that we should epistemically commit ourselves only to the mathematical or structural content of theories: it is this structural part of theories, the form of equations, that has regularly been retained over theory change, and it is thus this structural part that likely approximately gets right the structure – consisting in relations between objects and their features – of the metaphysical world. We are not warranted in any stronger epistemic commitment to physical theories.

Worrall discusses in this context how Fresnel's structure or form of the laws governing the relative intensities of reflected and refracted light beams in various circumstances expressed as equations, was retained through Maxwell's theory: although Fresnel's theory posited that light consisted of mechanical vibrations propagating in ether, its laws were *formally* similar to Maxwell's – which took light to consist of electrical disturbances propagating through vacuum – although the terms meant different things compared to the terms of Maxwell's equations. Using this case Worrall argues that with respect to the later Maxwellian picture, although Fresnel misidentified the nature of light and took it to consist in mechanical vibrations in an elastic ether, he nevertheless got the *structure* of light right. Stathis Psillos (1999) puts this view of Worrall's succinctly: "Mathematical equations which get retained in theory-change express real relations among entities for which we know nothing more than that they stand in these mathematically expressed relations to each other." (1999,143)

Interestingly, many including Worrall (1989, 2005) have interpreted Duhem as a structural realist. I consider Worrall in some detail later in this section, but to start with, consider Psillos:

> …it can be argued that Duhem's realism reaches up only to the structural level, so to speak. A natural classification is such that it gets right the *relations* among unobservable entities, but not necessarily the unobservable entities themselves. (1999, 35, emphasis as in original)

Milena Ivanova (2010) makes a very similar claim about Duhem. Barry Gower (2000) also takes Duhem to be a structural realist and I return to this below.

Are these views justified? Was Duhem in fact concerned with structure? We saw that structural realism as Worrall defends it has it that the structure of theory or theoretic structure – consisting in the forms of equations – gets (roughly) right the structure of the world or ontological structure – consisting in the relations between objects. The question

I ask then is, if similarly, a) what Duhem took to be approaching an ontological order is theoretic 'structure', and b) if this ontological order can be taken to be ontological 'structure' – both 'structures' being of the kind Worrall talks about.

To ask if Duhem was really talking about structure, let's start by answering a) above, i.e. looking at the theoretic level. As we have seen so far, what Duhem took to be approaching an ontological order was theoretic classification of experimental laws. Based on this, in a word, the answer to a) is no. Worrall's theoretic structure is mathematical structure consisting in the mathematical relations between terms of equations, whereas theoretic classification for Duhem was the grouping of disparate phenomena – via grouping of their experimental laws. Duhem meant 'classification' quite literally: he in fact builds this idea by treating it analogously to the way a naturalist classifies vertebrates into species. So by 'classification', Duhem had in mind grouping, ordering, and organizing of experimental laws (and therefore observable phenomena). This is evident from several passages like CL in the previous section. So Duhemian classification so far seems very different from Worrall's structure.

What facilitated/ engendered the theoretic classification of experimental laws were mathematical equations. Talking of a certain development in the study of light Duhem says,

> Light vibration will be the essential element by means of which the theory of light will be built; its components will serve in writing some *equations with partial derivatives and some boundary conditions*, condensing and

> classifying with admirable order and brevity all the laws of the propagation of light, its partial or total reflection, its refraction, and its diffraction." (1906/1954, 129, emphasis mine)

So what is *constitutive* of theory for Duhem is the mathematics[16] and what *comes out of it* is a condensed representation and classification of laws: physical theory "will have for its *object* ... the representation and natural classification of experimental laws, taken in a group." (107, emphasis mine) If Duhem had a positive metaphysical attitude towards this constitution, that would make him similar to Worrall. But according to the above account, Duhem was optimistic about *classification* tending to a natural order, but not about what is *doing* the classifying. So this greatly distinguishes him from Worrall.

Not saying anything about the reality of what is represented in the equations that do the classifying: their terms, their form, the boundary conditions they are subjected to etc., but saying only, very specifically, that the classification of phenomena brought about by these equations seems to reflect the groupings of the realities underlying the phenomena sounds like a strange position. But arguably, Duhem had underlying reasons for such a position. For him, the equations strictly *could not* have anything to do with reality, for they were mere *symbols.* Laws – both experimental and theoretical (which he called 'physical'), were for Duhem purely symbolic. They connect variables – symbols – that have a very complicated, theory-laden relation with reality that can only be spoken of in the context of theory. And he says that symbols, strictly speaking, cannot be true or

[16] See chapter I of part II in Duhem (1906/1954) for the role of mathematics in physical theory.

false, but only “more or less well selected to stand for the reality it represents” (1906/1954, 168)

But the situation is quite different when it comes to the classification that is borne out of theory. Although the classification is only meant to be a condensed representation of experimental laws – as above Duhem (in his more positivist moments) states that a classification can only be convenient or inconvenient but not true or false – one *can* meaningfully talk about a correspondence between the classification of laws and the true classification of realities behind the phenomena the laws are taken to govern. As has been discussed at length above, Duhem himself does this. It is difficult to talk of the counterparts of symbolic quantities like ‘pressure’ and ‘temperature’ in reality, and hence difficult to objectively talk of the relation between the two reflecting some ontological relation outside the context of theory. On the other hand, one can fairly easily talk of the (qualitative) groupings of the experimental laws, and thereby of the phenomena they describe, reflecting the affinity groupings of the underlying realities behind the phenomena – even if physics does not have access to this metaphysical realm. So while the laws of say optics and electricity are themselves symbolic, there isn’t any symbolism or theory-ladenness to saying that the laws of the two really (qualitatively) belong together in the sense that the realities behind light and electricity have some metaphysical affinity.

So Duhem does not seem to have been concerned with theoretic structure in the traditional structural realist sense. What may lead one to view him as a structural realist is the fact that Duhem talks of theory approaching an *ontological order* consisting in "affinities among the things themselves" (1906/1954, 26) and "the order in which a finished cosmology would *arrange the realities* of the world of matter" (1906/1954, 301, emphasis mine) So while I have (hopefully) shown that Duhem's theoretic classification is far from 'structure' as construed by structural realism, at the ontological level – addressing b) above – the Duhemian terms affinities/ groupings among unobservable things might sound similar to Worrall's structure: Worrall quotes Poincaré – on whom he bases his structural realism – as referring to (equations getting right) the *relations* between unobservable objects (1989, 118). So essentially, I'm asking this: Even though the classification of laws on the one hand and the mathematical form of equations on the other seem far removed from each other, might *what* the two are taken to ultimately approach – in the metaphysical world – be the same? But since we don't have direct access to the relations/ affinities in the metaphysical realm, we can only talk about them indirectly. And crucially, as we have seen, while the affinities in Duhem's metaphysical realm are reflected by the relations between experimental laws in a natural classification, the relations in Worrall's metaphysical world are those reflected by the structure of equations. So based on this again, Duhem does not seem to be concerned with structure in Worrall's sense.

There's another feature in Duhem that Worrall's view seems to share: retention of the mathematical machinery as physics progresses. Duhem talks of two parts of theory, one that attempts to explain (to reiterate, Duhem does not approve of this task for a theory, but does acknowledge the many attempts at explanation in the history of physics):

the explanatory part, and what he calls the ‘descriptive’ part, i.e. the other representative, classificatory part brought about by deductive-mathematical principles.. He asserts that the latter is not dependent on the former. The descriptive part of a theory is born and develops independently of the explanatory part “by the proper and autonomous methods of theoretical physics”, and “the explanatory part has come to this fully formed organism and attached itself like a parasite.” (1906/1954, 32). According to Duhem, theory owes its “power and fertility” entirely to the descriptive part. This part, which has a natural classification as its limit, constitutes “everything good in the theory” and is solely responsible for the power of theory to anticipate experience. “On the other hand,” says Duhem, “whatever is false in the theory and contradicted by the facts is found above all in the explanatory part; the physicist has brought error into it, led by his desire to take hold of realities.” (1906/1954, 32)

As an example, Duhem discusses Descartes’ law of refraction. He argues that while the law that posits a constant relation between the sine of the angle of incidence and the sine of the angle of refraction has stood the test of time and proved to be very successful in grouping together more fundamental laws and phenomena (like those of lenses and rainbows), Descartes’ attempts at the explanation of the nature of light – positing that it consisted in “rapid motions of incandescent bodies within an incompressible “subtle matter”” that is transmitted in it instantaneously – failed miserably.

This line is undoubtedly similar to one advanced by Worrall. A key aspect of Worrall's structural realism is that there is retention of structure across theory change. Gower (2000) in fact discusses this very argument of Duhem and takes him to be a structural realist based on it (86). But although Duhem says there is retention of the mathematical content across theory change, the main import of this retention was again the passing on of *classification of laws*:

> Thus, by virtue of a continuous tradition, each theory passes on to the one that follows it a share of the natural classification it was able to construct, as in certain ancient games each runner handed on the lighted torch to the courier ahead of him, and this continuous tradition assures a perpetuity of life and progress in science. (1906/1954, 33)

Importantly, Duhem does not say that the real metaphysical relations are reflected in the equations that classify laws. He in fact concludes this chapter with a reference to the *classification* of *experimental laws* becoming more and more natural (1906/1954, 54). So while Duhem, like Worrall, says that there is retention of mathematical content across theory change, he is not saying that what is retained directly reflects ontological order. It is rather the classification of experimental laws, brought about by this content, that reflects an ontological order. And recall that the content by itself simply couldn't have anything to do with reality for Duhem, for it is purely symbolic. So here again, his view is quite far from Worrall's.

One scholar who takes Duhem's positive attitude to be towards the classification of experimental laws – and nothing more than that – is Ernan McMullin (1990). He says about Duhem, "The only kind of realism that we can claim for him (and it is, of course, a crucial one for him) is that of the relationships he found in the laws of mechanics..."(427) Although McMullin once refers to Duhem's ontological order as "relational structure"(422), he doesn't attribute to Duhem the standard structural realist view that the mathematical structure of physical theory approaches an ontological structure[17].

Interestingly though, in a recent paper on the case of phlogiston in the history of chemistry, Ladyman (2011) advocates what seems to be a looser structural realism that lines up with the Duhemian view. In this paper Ladyman takes structure to be classification. He says of Epistemic Structural Realism (ESR) and Ontic Structural Realism (OSR), "ESR and OSR both involve commitment to the claim that science is progressive and cumulative and that the growth in our structural knowledge of the world goes beyond knowledge of empirical regularities." (2011, 97)

He argues that the above claim is supported by the phlogiston case:

> The empirical success of the theory was retained in subsequent chemistry since the latter agrees that combustion, calcification and respiration are all the same kind of reaction, and that this kind of reaction has an inverse reaction, and there is a cycle between plants and animals such that animals change the properties of the air in one way and plants in the opposite way. (2011, 99)

---

[17] It is possible though, that this is because this paper might have been written before Worrall's seminal 1989 paper that brought structural realism to the fore.

So he takes knowledge of the classification of the phenomena of combustion, calcification, and respiration together to be "structural" knowledge, and further says based on this, that we "identified a number of real patterns in nature" (2011, 100). For Ladyman then, the structure that a theory gets right need not be the quantitative relations expressed in the theory's equations. It could be a sorting of empirical phenomena/ laws into classes of the kind Duhem endorses.

Returning to Duhem, there are a few passages that could apparently lend themselves to a Worrallian interpretation of structure, but I argue that there is room to interpret them otherwise, in a way more consistent with Duhem's overall views.

Consider first,

> ...the more complete it becomes, the more we apprehend that the logical order in which theory orders experimental laws is the reflection of an ontological order, the more we suspect that *the relations it establishes among the data of observation* correspond to real relations among things, and the more we feel that theory tends to be a natural classification. (1906/1954, 26, emphasis mine)

While the first part of the quote talks of the relations between experimental laws, the second part talks about the relations between the "data of observation". This latter sounds like something that would find a place in the equations representing the experimental laws. For instance, it could mean the relation between the angle of incidence and the angle of refraction in a particular experiment with a lens. But "relations between data of observation" could also mean relations between data corresponding to observations of different phenomena. For instance, it could mean the common pattern that the data of the angles of incidence (i) and refraction (r) follow in relation to each other (sin i/ sin r = constant) in different cases, like in lenses and in raindrops. Note that here the relation in question is not that between 'i' and 'r', the terms of the equation, but rather between lenses and raindrops – or a law of lenses and a law of raindrops. This latter interpretation takes us back to the central Duhemian idea of theory ordering different phenomena (via ordering the experimental laws that describe them), and is hence the more plausible interpretation in my view. And of course, by "real relations among things", going by the rest of his arguments, Duhem must mean the real affinities or groupings among things.

Next, consider:

> If on the contrary, we recognize in the theory a natural classification, if we feel its *principles* express profound and real relations among things, we shall not be surprised to see its consequences anticipating experience and stimulating the discovery of new laws (1906/1954, 28, emphasis mine)

Note that Duhem is talking of the *principles* of theory reflecting real relations. In using the term principles, Duhem seems to be saying that the relations expressed in the *constitution* of theory are getting closer to the real relations. But given the bulk of his arguments about the *classification of laws* reflecting an ontological order – consisting in the groupings between things – this line can be interpreted as meaning that the principles express, *via the classification they bring about*, the real relations, i.e. groupings or affinities among things – rather than that the principles express the real relations in their (mathematical) statements.

Next,

> So the physicist asserts that the order in which he arranges *mathematical symbols in order to constitute a physical theory* is a clearer and clearer reflection of an ontological order according to which inanimate things are classified. (1906/1954, 299, emphasis mine)

It's not clear what Duhem is referring to here by "symbols": ones within a theoretical-mathematical equation – its terms – that contribute to classifying experimental laws, or whole laws themselves – for an experimental law, for Duhem, was also itself a symbolic representation of phenomena. If it is the latter, and Duhem is referring to whole laws as "mathematical symbols", then he is again just talking of the order among laws. But if it is the former, and he's talking about symbols within equations, then the "order" here could be seen as structure in Worrall's sense. But going by the bulk of his views, it's very likely that Duhem means (experimental) laws by 'symbols' here.

Finally, there is one remark that sticks out of Duhem's discussion of natural classification where it admittedly sounds like he's talking about Worrall-type structure. In fact, Worrall (2005, 4) himself takes Duhem to be a structural realist based on this very remark:

> **S**: The highest test, therefore, of our holding a classification as a natural one is to ask it to indicate in advance things which the future alone will reveal. And when the experiment is made and confirms the predictions obtained from our theory, we feel strengthened in our conviction that the relations established by our reason among abstract notions truly correspond to relations among things. (1906/1954, 28)

This above quote by Duhem is also seen in Worrall (1989).

The part of Duhem's quote that I want to draw attention to is the following: "we feel strengthened in our conviction that the relations established by our reason among abstract notions truly correspond to relations among things". What are abstract notions? For Duhem, abstraction is a process by which the mind analyzes a multitude of independent, concrete facts, takes what is general and common to them and summarizes them in an experimental law. (1906/1954, 22) A law is a "general proposition tying together abstract notions". (1906/1954, 55) Further, a theoretical/ physical law that condenses a number of experimental laws introduces another level of abstraction. (1906/1954, 166-167) For Duhem, 'pressure', 'electromotive force' and the like are examples of abstract notions: they symbolize certain properties of observable phenomena and are "related to concrete realities only by long and complicated intermediaries". (1906/1954, 148) So what the above quote seems to be saying is that we feel that the relations that a law establishes in an equation between these abstract notions – these abstract, symbolic quantities – reflect real relations between things. This sounds like Worrall-type mathematical structure.

What to make of Duhem's views now? On the one hand, there is overwhelming evidence of him talking of groupings of laws by theory getting to an ontological order in remarks like “The groups it establishes permit hints as to the real affinities of things.” (1906/1954, 30), and that theories “group appearances in the same way realities are grouped” (1906/1954, 31). We saw this in detail earlier. But on the other hand, we have the statement under discussion, about the relations between abstract notions corresponding to the real relations between things. This surely sounds different from and irreconcilable with the bulk of Duhem’s views on classification.

6.2.3. Conclusion

Because of the last four quotes, particularly the last one, it is more ambiguous now as to what Duhem took as felt by the physicist to be progressively getting right the real relations among underlying realities: the relations between experimental laws, or the relations expressed in the equations. Let’s call classification in the first sense – that of experimental laws – the ‘standard view’ of Duhemian classification for it is this view that Duhem espouses in most of the book. According to the standard view of Duhemian classification then, his classification cannot be seen as structure in the Worrallian sense. But as we saw, if structure is taken to be the classification of laws/ phenomena as Ladyman takes it to be, then Duhem can be taken to be concerned with structure.

But according to the latter view of classification above – especially supported by the last quote above (S) – Duhem seems to have been concerned with relations within equations which very much sounds like Worrallian structure. If this is what Duhem meant, surely, the two theoretic relations in question – within equations, and of laws –

cannot be equivalent in any straightforward sense. Multiple sets of theoretical equations can in fact bring about the same classification scheme and multiple laws can bring together the same set of facts. The connection between the two kinds of classification is thus not clear, and Duhem does not explain.

So what was the object of Duhem's positive attitude towards physical theory? The answer most likely is that it was only the classification of experimental laws brought about by theory (in the sense that he thought that we are compelled to believe that physical theory is progressively getting something right about a metaphysical order). The less likely answer is that it was structure, as in the form of equations of theory, the order in which it arranges symbols. Given this ambiguity, was Duhem being inconsistent with regard to the object of his realist attitude? Probably. But on a more charitable reading, we could say that at the least, may be Duhem was not being inconsistent – may be there *is* a connection between the classification of laws and the mathematical form of equations – but D*uhem did not make the connection*: he did not bridge the gap between the two "structures".

6.3 The Rationale behind Pierre Duhem's Natural Classification

6.3.1 Background

Traditional accounts have pictured Pierre Duhem as a paradigm instrumentalist: they take it that he maintained that successful theories of physics do not tell us how nature operates, but are just convenient tools that 'save the phenomena' and represent and classify empirical laws. Recently though, there has been a spate of work centered around his idea of natural classification, maintaining that Duhem was a realist of sorts – maybe a no-miracles type[18] realist, or a plausibility realist.

Here I argue that none of these accounts are satisfactory. I contend that Duhem had a positive attitude to physical theory – especially to what he calls 'logically unified' theory – which arguably may be seen as a realist attitude. But as the traditional accounts go, he was by no means a scientific realist in the usual sense: he did not think physical theories made – or approached or approximated – true claims about underlying reality. Duhem was not a no-miracles type realist or a plausibility realist. I shall argue that the passages that are claimed in recent work to support these different realist readings of Duhem do nothing of the sort. Contrary to the underlying supposition of these new interpretations, Duhem does not ground his pro-attitude to logically unified theory in the past predictive success of the theory. Rather, he thought the physicist 'feels' or 'surmises' that theory affords a classification of experimental laws that progressively reflects a metaphysically true classification of things. Beyond this, Duhem is concerned not directly with the truth, in some sense or another, of the claims of successful theories in physics but instead with the pragmatic rationality of the physicist. By affirming that the

[18] What is claimed is not that he was a scientific realist, but rather that he was a no-miracles type structural realist, like Worrall (1989): a no-miracles type argument for structural realism goes that it would be a miracle if our theories don't get the structures of reality (approximately) right. As argued in the previous section, it is not clear if Duhem was concerned with structure, but it is clear that that object was classification of laws. Here I'm interested in his rationale behind the realist attitude.

classification of laws afforded by theory approaches a "natural" classification, 1) The physicist who bets on the side of predictive success avoids folly and exposure to ridicule in doing so; and 2) The physicist who pursues logically unified theories rationally justifies the pursuit by identifying a meaning or purpose to the pursuit. But again, contrary to the traditional instrumentalist accounts I think this does warrant a realist reading of his attitude to theory. But importantly as above, the rationale behind sustaining this realist attitude was not that historical evidence compels us to hold it, as the no-miracles camps claim, but rather was a pragmatist one: *it rationalized, on pragmatic grounds, the physicist's activity of pursuing theory*. (Hence if the rationale behind a philosophical position is taken as the basis for labeling the position, this view might be seen as a form of instrumentalism.) Karen Darling's (2003) motivational realist reading of Duhem comes closest to what I take to be the right interpretation of him among existing views, but I propose here that Hasok Chang's (2009) ideas on the intelligibility of activities and pragmatic rationality fit Duhem even better.

6.3.2 Introduction to Duhemian Natural Classification

A physical theory for Duhem represents, organizes and classifies experimental laws. Duhem discusses the case of "light vibration" for instance: he says it is given a direction, an intensity, and is geometrically represented by a line with a periodically varying length, and the components of light vibration will "will serve in writing some equations with partial derivatives and some boundary conditions, condensing and classifying with admirable order and brevity all the laws of the propagation of light, its partial or total reflection, its refraction, and its diffraction." (1906/1954, 129) Duhem held that a theory should also be logically unified in that it should not employ multiple, incompatible ways of classifying laws.

Duhem rejects the view that theories go beyond this and explain phenomena in terms of underlying causes, for such explanation, for him, is the business of metaphysics. Physics for him is only concerned with studying phenomena, finding experimental laws, and organizing and classifying these laws. According to Duhem, it is beyond the means of physics to study the causes underlying the phenomena: the experimental method of physics does not have the resources to provide any positive metaphysical knowledge.

Further, Duhem argues that given the vast number of metaphysical disagreements and irreconcilable metaphysical positions (regarding the nature of matter, of light, of magnetism etc.) throughout the history of physics, the physicist should not get involved with such concerns. Physics, according to Duhem, must be entirely separate from metaphysics in its interests and concerns. He cites the perennial rampant disagreements between different metaphysical schools and argues that pursuing explanations would make physics subordinate to metaphysics; whereas physics, according to him should be an autonomous pursuit. Hence for him, metaphysical explanations have no place in physics.

For Duhem, physical theories provide an "economy of thought" and serve to store an otherwise intractable number of experimental laws – laws of the kind we discover and record in experiment and careful observation – in “condensed representations” (1906/1954, 23). But “it matters little” for him whether the operations performed to combine various hypotheses together “do or do not correspond to real or conceivable physical transformations”. (1906/1954, 20) All that matters is that theories must be consistent with the laws they represent.

Three quotes from Duhem are illuminating when looked at together. First, in light of the above claims about theory being disconnected from metaphysical reality, Duhem allows a very limited criterion for calling a theory *true*:

D1: “Agreement with experiment is the sole criterion of truth for a physical theory.” (1906/1954, 21)

As above, in addition to effective representation, what is most important about theories for Duhem is that they *classify* experimental laws. Duhem says about theory,

> Thus alongside the laws which govern the spectrum formed by a prism it arranges the laws governing the colors of the rainbow; but the laws according to which the colors of Newton's rings are ordered go elsewhere to join the laws of fringes discovered by Young and Fresnel; still in another category, the elegant coloration analyzed by Grimaldi is considered related to the diffraction spectra produced by Fraunhofer. The laws of all these phenomena, whose striking colors lead to their confusion in the eyes of the simple observer are, thanks to the efforts of the theorist, classified and ordered. (1906/1954, 24)

So Duhem means ‘classification’ quite literally: what he has in mind is just grouping and ordering. Is this classification in any way indicative of any features of a metaphysical reality? Not in any straightforward sense, for Duhem. Duhem says, “… physical theories are only a means of classifying and bringing together the approximate laws to which experiments are subject; theories, therefore, cannot modify the nature of these experimental laws and cannot confer absolute truth on them.” (1906/1954, 171) This line is again echoed in *Essays in the History and Philosophy of Science* (1996). He says, “Laws of physics retain exactly the same sense when a theory connects them as when they are dispersed and isolated.” (Duhem, 1996, p.36) Theorizing – i.e. representing and classifying laws economically – has no effect on the character or content of physical science: “It remains physics; it does not become metaphysics.” (1906/1954, 36) He adds,

D2: “A classification, in fact, is not a judgement. It can be convenient or inconvenient, good or bad; it cannot be true or false.” (1906/1954, 37)

Despite D2, Duhem greatly valued the classifying power of theory. Not only did he think this classification was 1) “beautiful” (1906/1954, 24), but also that 2) the elegance and efficacy of the classification persuade us to believe that it tends to reflect a *natural, underlying classification or order*; to believe that the relations among phenomena established by theory “truly correspond to relations among things” (1906/1954, 28). A natural classification for Duhem is a limiting case of regular theoretic classification: a natural classification is a perfect, ideal classification of all experimental laws, and it is *natural* because it would perfectly mirror ontological relations between the realities behind the phenomena.

So in what seems like a stark contrast to D2, in the following remark on physical theory – and several such remarks throughout *Aim and Structure of Physical Theory* (A&S) – that is especially telling of his anti-instrumentalist dispositions, Duhem says,

D3*:* "…we feel that the groupings established by our theory correspond to real affinities among the things themselves." (1906/1954, 26)

(This affinities-speak has motivated a structural realist reading of Duhem, which I discussed in the previous section. Here, I'm interested in the rationale behind this "feeling".) Importantly, Duhem says that the physicist cannot account for this conviction through the "method at his disposal", which is "limited to the data of observation". "It therefore cannot prove that the order established among experimental laws reflects an order transcending experience…" (1906/1954, 27) But he continues, "...while the physicist is powerless to justify this conviction, he is nonetheless powerless to rid his reason of it…. He cannot compel himself to believe that a system capable of ordering so simply and so easily a vast number of laws, so disparate at the first encounter, should be a purely artificial system." (1906/1954, 27)

Duhem expresses this idea of the physicist's intuition about theory progressively reflecting a natural classification - an ontological order - in several parts of A&S. Consider for instance,

> … [The physicist] will note that physical theory in its successive advances tends to arrange experimental laws in an order more and more analogous to the transcendent order according to which realities are classified, that as a result physical theory advances gradually toward its limiting form, namely, that of a *natural classification*... (1906/1954, 297)

Here then is a succinct version of what Duhem thought the physicist feels about physical theory. I shall call it the *thesis of natural classification (TNC)*:

TNC: *As physical theory advances it progressively approaches a 'natural classification' of experimental laws, i.e. one that reflects a metaphysical order.*

(There was an important connection for Duhem between TNC and the logical unification of theories, which I discuss in detail later.)

Duhem asserted TNC based on the 'feeling' he got from looking at the historical development of physics, and the working attitude of a typical physicist who took natural classification to be the ultimate ideal of physical theory. So Duhem seems to have had a positive ontological attitude toward advancing physical theory in the sense that he thought the physicist cannot "rid his reason" of the conviction that the classification of laws afforded by it progressively approaches a natural classification that reflects an underlying, metaphysical order although it is beyond the scope of physics and logic to justify such a belief.

6.3.3 Duhem's Rationale : Existing views centered on the Success of Novel Predictions

What was the basis for TNC? I now discuss some of the existing views on this. A most popular bit from Duhem that appears in many discussions of his natural classification is what he had to say on the discovery of new, as yet unknown laws. Although what Duhem says on this is important to the discussion, I think the importance has been exaggerated and this has led to some misinterpretations. Let's first look at what Duhem had to say:

> **P**: If the theory is a purely artificial system, if we see in the hypotheses on which it rests statements skillfully worked out so that they represent the experimental laws already known, but if the theory fails to hint at any reflection of the real relations among invisible realities, we shall think that such a theory will fail to confirm a new law. That, in the space left free among the drawers adjusted for other laws, the hitherto unknown law should find a drawer already made into which it may be fitted exactly would be a marvelous feat of chance. It would be folly for us to risk a bet on this sort of expectation. If on the contrary, we recognize in the theory a natural classification, if we feel its principles express profound and real relations among things, we shall not be surprised to see its consequences anticipating experience and stimulating the discovery of new laws; we shall bet fearlessly in its favor.
>
> The highest test, therefore, of our holding a classification as a natural one is to ask it to indicate in advance things the future alone will reveal. And when the experiment is made and confirms the predictions obtained from our theory, we feel strengthened in our conviction that the relations established by our reason among abstract notions truly correspond to relations among things. (1906/1954, 28)

This may prima facie strike one as being similar to the traditional argument for scientific realism: the no-miracles argument (NMA) first put forward by Hilary Putnam (1975), a species of Inference to the Best Explanation (IBE). The argument goes roughly that the best explanation we can infer from the vast predictive successes of a scientific theory is that that theory is true. (There has been quite a vast variance in understanding 'truth' here, but most realists agree that truth consists in some kind of 'latching on' of theory to underlying reality.) Otherwise, their empirical success would be a miracle, a "marvelous feat of chance". In fact, Andrew Lugg (1990) says just this - that Duhem espoused a version of the NMA.

But I think this is a misinterpretation for it is mistaken at two levels. Duhem's argument is by no means an IBE. There's certainly no 'I', nor is there an 'E'. Duhem is not saying that we infer TNC: TNC is an intuition, a feeling, not the result of an *inference* (based on success, or anything else). Secondly and more importantly, Duhem is not concerned with explanation of success at all, which is what the NMA argument is centrally about. He's not saying that the physicist *explains* success of theories with TNC[19]. When he says that it would be a marvelous feat of chance if a purely artificial theory made a right prediction, unlike realists, Duhem is not making a (direct) leap from that view to the conclusion that the theory is (therefore) approaching a natural classification, in a bid to explain the successful predictions it's made. Rather, as in P

[19] It is important to note here that Duhem *was* concerned with past successes, albeit in a different context: he was concerned with the value of recognizing historical continuity in the evolution of physical theory. He believed that the physicist must study past theories to understand the evolution and direction of progress in order to get hints about its future path. But Duhem makes no claim about this continuity indicating any kind of truth of (any components of) theory.

above, what he's saying is that it would be "folly" to bet in favor of a prediction made by such an artificial theory. What exactly does Duhem mean by folly? I think he means it in the sense or pragmatic prudence. When one is in the business of making (good) predictions based on a theory, it is only beneficial to her if she takes her theory to be such that it is capable of making good predictions. If she both assumes that the theory is not reliable *and* makes predictions based on it, there is a lack of internal coherence – between her beliefs and her goals. In the present context of Duhem, of course reliability of the theory is taken to consist in natural classification.

Here then is my central point: Duhem is here concerned primarily with the pragmatic rationality of the physicist – in the sense of what would and wouldn't count as folly or unreasonable with respect to pursuing theory, and the rationale behind that – but not with the success of theories. This is the crux of my argument and I return to this in more detail in a bit.

Unlike Lugg, John Worrall (2011) thinks that Duhem is not advancing an IBE. Although it cannot be called a no-miracles *argument* per se, it is still, he thinks, a no-miracles *intuition*. Worrall starts with the classic question posed by the realist to the antirealist: "How on earth…could a theory score a dramatic predictive success… unless its claims about a reality "underlying" the phenomena… are at least approximately in tune with the real underlying structure of the universe?" (2011, 11) He then quotes the last two sentences of P in support. Based on this, Worrall takes Duhem to be addressing the question posed above - he takes him to be endorsing the "realist-leaning force of predictive successes…" (2011, 12)

Stathis Psillos (1999) attributes to Duhem a version of realism he calls plausibility realism. Psillos holds that even if, as Duhem thought, the apprehension about TNC may be outside the scope of science, it is nonetheless a *philosophical* argument and is not without "rational force". In his words, "Plausibility is certainly a reason to prefer one position over another." (1999, 36) Psillos starts by explaining arguments of J.J.C Smart that seem to have been a precursor to the NMA. For Smart, if theories were purely instrumental, then it would be very odd that the phenomena of the world would be such as to make such a theory true of them. On the other hand, if the theory is viewed in a realist way, there would be no such "cosmic coincidence". Psillos points out that Smart's argument, at first glance, seems to simply be a variant of NMA/ IBE, but that on a careful look, one can see that "Smart's argument is not meant to be an inference to the best explanation. It is more of a general philosophical argument, what is sometimes called a plausibility argument." (1999, 71).

For Smart, says Psillos, "the argument for realism is largely a priori" (1999, 71) and intuitive: the conclusion of the argument, even though not logically compelling, would be intuitively plausible, persuasive and rational – not because the argument can be recognized as a trusted inferential scheme, but because of intuitive plausibility considerations. Psillos then brings to our attention that Grover Maxwell also advanced a similar argument based on the success of science. Psillos thinks that Duhem fits this picture as well and is, along with Smart and Maxwell, a plausibility realist. It looks like Psillos is taking off from where Worrall stopped: Worrall took TNC to be a no-miracles intuition, and Psillos is defending a more rational basis for that intuition on plausibility considerations. At first glance, it may seem that Smart's arguments are strikingly similar to Duhem's, as Psillos says.

But I contend that both Worrall and Psillos are mistaken in their interpretation of Duhem. As above, exploring the source of the empirical success of theories is not Duhem's main concern. I think Duhem was neither concerned with explaining success, nor with the plausibility or implausibility of the success. The central subject matter of Duhem's discussion is the pragmatic rationality of the physicist. I shall now try to develop this argument more fully.

6.3.4 The TNC rationale – an initial response, and its distance from the success of novel predictions

Let's examine P above closely. First, notice that the successful novel prediction in question only reinforces or strengthens an intuition we already had in the first place: for Duhem, the intuition is not based on this success as Worrall/ Psillos make it out to be. Second, nor is it the case for Duhem that the intuition is based on past successes as Worrall and Psillos hold, and gets strengthened by "one more" success. Importantly, Duhem is not saying that we intuit TNC based on success. This is what mainly distinguishes Duhem from the no-miracles intuition/ argument camp.

What then is the TNC intuition based on? There is to be found in Duhem, a cursory and a deeper version of a response to this question. I submit that both are crucial to showing that Duhem does not fit the above standard realist interpretations. In the early discussions of natural classification in A&S*,* we see that for Duhem, the TNC intuition is rooted in the *structuring* of experimental laws that physical theory provides - its ability to order, organize, and unify the laws, *not* its empirical success. Call this Duhem's "argument from structure" – or, keeping with Worrall's distinction, the "intuition from structure" (as opposed to argument/ intuition from *success*).

Talking of a certain theory of vibration of light, Duhem says,

> … when, after much groping, we succeed in formulating with the aid of this vibration a body of fundamental hypotheses, when we see in the plan drawn by these hypotheses a vast domain of optics, hitherto encumbered by so many details in so confused a way, become ordered and organized, it is impossible for us to believe that this order and organization are not the reflected image of a real order and organization (1906/1954, 26)

He continues,

> ...the more complete it becomes, the more we apprehend that the logical order in which theory orders experimental laws is the reflection of an ontological order, the more we suspect that the relations it establishes among the data of observation correspond to real relations among things, and the more we feel that theory tends to be a natural classification. (1906/1954, 26)

So Duhem's point is that the physicist cannot resist the feeling or intuition that the classification that theory provides is approaching a natural classification, looking at the elegance and efficacy of the classification of laws it affords. *So the TNC intuition is rooted in the structure that theory provides for empirical laws, not in the empirical success of the theory.* Empirical success was only incidental to this structure. In fact, Psillos' idea of an intuitive, plausibility-based rationale can fit Duhem: if we replace past predictive successes with the structuring of laws. We can reasonably claim that based on theory's structuring of empirical laws – and not based on past predictive success – it is intuitively implausible, according to Duhem, that theory is not progressively reflecting a natural classification. This is the initial idea of the rationale behind natural classification presented by Duhem in A&S. I discuss below the deeper and more elaborate version of Duhem's views on the basis of the TNC intuition but before that I wish to make a final, important point on what Duhem had to say on novel predictive success, as this has been a bone of interpretational contention.

I want to point out that in A&S, this discussion on the intuition based on the order/ structure theory provides for experimental laws *precedes* P quoted above. It is after this part on the 'intuition from structure' that Duhem mentions the discovery of new laws. And he says, "There is *one circumstance which shows with particular clarity* our belief in the natural character of a theoretical classification…"(1906/1954, 27, emphasis mine) – and this circumstance is that of a theory predicting a new law. Duhem asks, "Now, on the occasion when we confront the predictions of the theory with reality[20], suppose we have to bet for or against the theory; on which side shall we lay our wager? " (1906/1954, 28) It is here, right after this, that P appears. So: the success of a novel

[20] Here Duhem means experimental reality, not some underlying reality.

prediction is not the basis for the TNC intuition, it only reinforces/ vindicates it with "particular clarity". This starkly sets Duhem apart from the no-miracles camps.

6.3.5 The rationale behind TNC elaborated – making sense of pursuing physical theory

TNC is intuited based on the classifying of laws by theory, but why uphold the intuition? This is a very significant question for Duhem. His answer is that pursuing physical theory would be justified – in that it would a worthwhile activity for the physicist – only if TNC is affirmed. Without TNC, pursuing theory would be meaningless and unjustified. The ultimate answer to the question, '*Why* pursue physical theory?' for Duhem lay in affirming TNC, for it gave the physicist good reasons to pursue theory. At the core of pursuing physical theory for Duhem, were 1. making novel empirical predictions with confidence, and 2. striving for *logical unity.* And I show below that according to Duhem, for both of these to be justified for the physicist – in the sense explained above – she has to affirm TNC. I now discuss these in turn.

The physicist's confidence in novel prediction[21]

[21] It is probably worth noting that most 20th century philosophers meant by "novel prediction" the prediction of new *facts*, but Duhem meant both prediction of new facts and of new *laws*, and provides many examples of both – for instance, Poisson's prediction based on diffraction, of the nature of the image on a screen behind a small, circular and opaque screen intercepting light from a point source (1906/1954, 28); and Huygens' prediction of the laws governing the two refracted rays (ordinary and extraordinary) of a single light ray traveling from air into a crystalline medium (1906/1954, 35). It should also perhaps be mentioned that (unfortunately) he switches back and forth between the two ideas and examples of each without noting the difference.

With respect to the pursuit of novel predictive success, what Duhem is interested in, I contend, is the pragmatic rationality of the physicist – making sense of the chosen activity of prediction – and not explaining, or defending the (im)plausibility of, the success. As above, Duhem starts by asking which side we might take with respect to the question, 'Is the theory going to make a correct prediction?' As he makes clear later in A&S (quoted below), Duhem contends that no physicist is agnostic or pessimistic about this: physicists are routinely committed to the pursuit of novel predictions and engage in them with confidence. Given this commitment, how to justify the confidence, when physics does not have any resources to justify it? Duhem says that if we take the position that the theory is in fact going to make a correct prediction, while at the same time denying that the theory "hints at any reflection of the real relations among invisible realities" – i.e. denying TNC, then we would be engaging in folly since physics does not have any resources to justify this confidence. But if TNC is affirmed, then being confident about a novel prediction doesn't seem unreasonable at all – the physicist is rationally justified in the sense of being protected from ridicule. Hence, TNC cannot be denied and needs to be affirmed.

This argument is quite easily gleaned from the following discussion that appears in the appendix. Concerning the physicist's confidence in the novel empirical prediction of a theory, Duhem wonders, despite the physical method not lending any support to this confidence,

> In what physicist do we ever meet such perfect indifference concerning the result of a test…? The physicist knows quite well that strict logic absolutely allows him only this indifference and that it authorizes no hope of agreement between theoretical prophecy and the facts; nevertheless, he waits for this agreement, counts on it, and regards it as more probable than the refutation. (1906/1954, 298)

He continues,

> **C**: None of the rules governing the handling of experimental method justify this confidence in the theory's foreknowledge, and yet this confidence does not seem ridiculous to us. Furthermore, if we harbored some intention to condemn its presumption, the history of physics would surely not take long to compel us to modify our judgment; indeed, it would cite innumerable circumstances in which experiment confirmed down to the smallest details the most surprising predictions of theory.
>
> Why then can the physicist, without exposing himself to ridicule, assert that experiment will disclose a certain law because his theory demands the reality of this law…?… Obviously because… in the physicist's theory there is something like a transparent reflection of an ontological order. (1906/1954, 298)

So again, Duhem is not saying that owing to the novel predictive success of the theory, we feel that it approaches a natural classification. He is rather saying that this feeling that theory approaches a natural classification is what makes the physicist's confidence in a given theoretical prediction not ridiculous. These are two separate arguments.

To be sure, a Worrall/ Psillos type no-miracles argument/ intuition could imply Duhem's position. If you look at a theory's track record of novel predictive success, and then say based on that that the theory is likely to be true, then that inference to truth might rationalize your confidence about the theory's next novel prediction. So Duhem's argument seems to be consistent with an NMA, but my point is that Duhem is simply not making that argument.

Although physics and logic have no resources to justify the confidence in a prediction, the path determined by history preconditions the construction of theory and lends support to the confidence the physicist has in it: "so the history of physics lets us suspect a few traits of the ideal theory to which scientific progress tends" (1906/1954, 303). But although our confidence in a prediction is backed by the history of success, this history itself is not comprehensible, and physicists' pursuit of theory is not legitimized, unless TNC is asserted:

> Physical method is powerless to prove this assertion is warranted, but if it were not, the tendency which directs the development of physics would remain incomprehensible. Thus in order to find the title to establish its legitimacy, physical theory has to demand it of metaphysics. (1906/1954, 298)

So TNC is what legitimizes physical theory and justifies its pursuit. Note that the sense of justification here is not either that the history of predictive success or the successful structuring of empirical laws supports the truth of the claim expressing the new prediction. Rather, that physicists in affirming TNC are justified in their confidently engaging in novel prediction in the sense that they are not exposing themselves to ridicule in doing so. I shall return to this justification in a little more detail in a bit. Duhem's overarching point is that if we worked to give up the TNC intuition, we would have to be completely indifferent towards the outcome of a novel prediction (failing some other good reason, which does not seem available). He then points out that no physicist has this complete indifference, and history in fact encourages her to be confident. The question is then, what explains this historical trend and justifies this optimism that physicists have had for years? The answer is TNC. Hence, denying TNC would render the physicist's activity of making theoretical predictions with confidence unjustified. And then, when theory confirms the prediction we bet in favor of, our conviction in TNC becomes stronger: we feel "strengthened in our conviction that the relations established by our reason among abstract notions truly correspond to relations among things" (1906/1954, 28)

So the task at hand for Duhem is identifying what justifies the decision on which way to bet with respect to the prediction of a theory – it is the TNC intuition. The task is not explaining/ exploring the (im)plausibility of the theory's success.

Unity in Physical Theory

Recall the intuition from structure. I claimed that according to Duhem the physicist is led to intuit from the great order that theory provides for experimental laws, that this order must tend to a natural classification. For Duhem, this order comes from the physicist's efforts to construct *logically unified* theories rather than vastly disparate ones. By a logically unified theory, Duhem means a theory that represents "the whole group of natural laws by a single system all of whose parts are logically compatible with one another." (1906/1954, 293) This aspiration for unity (AfU) is a central idea in Duhem's arguments: “Every physicist naturally aspires to the unity of science.” (1906/1954, 103) So Duhem thought that every physicist’s mind is naturally predisposed to work toward the unity of science. In a similar vein, speaking of those “working for the advance of physics”, Duhem asks, “Is there a single one among them who hesitates for an instant to prefer a rigorously coordinated theory to a junk heap of irreconcilable theories…?” (1906/1954, 295) Duhem explains that to a staunchly logical mind, to one that strictly confines itself to the methods of physics, this incoherent heap should pose no problem: all that such a physicist will be concerned with is consistency between theory and experiment and not mixing up incompatible theories (1906/1954, 294). Also, the law of economy also does not demand logical unity since different kinds of minds construe economy differently. The “ample” minds according to Duhem will “find that the logical labor of coordinating diverse fragments of theory into a single system is considerable… They will not by any means judge the passage from incoherence to unity an economical intellectual operation.” (1906/1954, 102) Therefore, he says, “Neither the principle of contradiction nor the law of economy of thought permits us to prove in an irrefutable manner that a physical theory should be logically coordinated…” (1906/1954, 102)

Further,

> If we limit ourselves to invoking merely the grounds of pure logic[22], of that logic which allows us to determine the object and structure of physical

[22] It may seem like Duhem is being inconsistent when he says that a) *logic* does not condemn the use of two incompatible or contradictory theories, and b) the physicist wants theories to be *logically* unified. But Duhem simply means that logic does not forbid using two logically incompatible theories *as long as their premises are not mixed up* and as

> theory, it is impossible to… condemn a physicist who would claim to represent by several logically incompatible theories either diverse sets of experimental laws or even a single group of laws… (ibid., p.294)

Despite this, he asserts that physicists would not give up this pursuit of unity:

> they feel that this way is the right one; they have an intuition that logical unity is imposed on physical theory as an ideal to which it tends constantly; they feel that any lack of logic, any incoherence in this theory, is a blemish, and that the progress of science should gradually remove this blemish. (1906/1954, 294)

And:

> The physicist, then, finds himself in an irresistable aspiration toward a physical theory which would represent all experimental laws by means of a system with perfect logical unity… (1906/1954, 296)

Crucially for Duhem, TNC goes hand in hand with this drive for unity. Perfect logical unity is an important mark of a theory that yields a natural classification. According to Duhem, physicists "have an intuition that logical unity is imposed on physical theory as an ideal to which it tends constantly…" (1906/1954, 294) And, they affirm that "logical unity is a characteristic without which physical theory cannot claim this rank of a natural classification." (1906/1954, 296)

---

long as they are each used separately to order and organize laws. Yes, they *are logically* incompatible and contradictory when looked at together, but the physicist is not committing a contradiction as long as she doesn't mix them up. Logical unity – which is itself not demanded by logic – then, is about avoiding such contradictory theories and constructing theories that are logically unified.

To understand Duhem's idea of unity better, here's a revealing quote. In a negative discussion of unity – while arguing that logic does not demand such unity – Duhem says,

> …how can we draw from the code of logic the right to condemn a physicist who employs, for the sake of ordering laws, different methods of classification, or a physicist who proposes, for the same set of laws, diverse classifications resulting from different methods? Does logical classification forbid naturalists to classify one group of animals according to the structure of the nervous system, and another group according to the structure of the circulatory system?....Thus a physicist will logically have the right first to regard matter as continuous and then to consider it as formed of separate atoms, to explain capillary phenomena by forces of attraction acting between stationary particles, and then to endow these same particles with rapid motion in order to explain heat phenomena. (1906/1954, 101)

Duhem concludes by saying that even though unity of theories is not demanded by strict logic, those very physicists who have come up with theories whose "various parts cannot be fitted together" have only done so regretfully, and that everybody naturally aspires for logical unity.

Duhem then contends that if physical theory were just a tool, an instrument, then it could only be "convenient or inconvenient" (ibid., p.334), and the physicist would not be intuitively compelled to see in the logical unity of theories, a natural order. But, Duhem says,

> **M**: When the physicist after submitting his science to this careful examination returns into himself, and when he becomes aware of the course of his reasoning, he at once recognizes that all his most powerful

> and deepest aspirations have been disappointed by the despairing results of his analysis. No, he cannot make up his mind to see in physical theory merely a set of practical procedures and a rack filled with tools. No, he cannot believe that it merely classifies information accumulated by empirical science without transforming in any way the nature of these facts or without impressing on them a character which experiment alone would not have engraved on it. If there were in physical theory only what his own criticism made him discover in it, he would stop devoting his time and efforts to a work of meager importance. (1906/1954, 334)

Duhem concludes that the physicist is compelled to affirm that physical theory confers a certain knowledge of the external world on the physicist that is irreducible to merely empirical knowledge – this knowledge is that of a metaphysical order of nature, which is reflected by a natural classification, towards which theory's classification of laws tends. That is, Duhem is reiterating TNC here. Based on the above quote, was the TNC intuition for Duhem *motivational*? Does the physicist assert TNC because it motivates the pursuit of physical theory? Based on Arthur Fine's characterization of Einstein's philosophy as "motivational realism", Karen Darling (2003) proposes that the same is applicable to Duhem. Speaking of the dual intuitions of the AfU and TNC, she says, "Duhem identifies in the physicist two realist intuitions that, as reasons of the heart are unable to ground a pro-realist argument, although they do, he claims, provide motivation for the practice of science." (2003, 1134)

She explains that according to Fine, Einstein adopted motivational realism with "no deliberate intention or program, but straight from the heart" and that it reflected "the dues that Einstein felt worth paying for his passionate commitment to science, and for the meaning that scientific work gave to his life" (2003, 18) Darling tells us that motivational realism according to Fine has the following three key features:

1. It is a realist attitude.
2. It motivates and gives meaning to scientific activity.
3. It does not espouse a global doctrine or a specific set of beliefs about reality. (2003, 1134)

Motivational realism also has it that one cannot meaningfully talk of the truth of theories where some correspondence theory of truth is assumed. According to Fine, Einstein did not think that theories were getting to some underlying metaphysical truth. Realism was not a set of beliefs about the relationship between theory and reality, but was rather a "program" or attitude motivating scientific activity. Darling argues that Duhem too had a realist attitude – based on Duhem's remarks such as "…it is impossible for us to believe that this order and organization are not the reflected image of a real order and organization" (Duhem, 1906/1954, 26) above. Given that I have denied the traditional realist readings of Duhem, it is important to look into this point. Did Duhem have a realist attitude? I think it is indeed fair to say he did, based on several remarks like the one above. For instance recall that he says that when the prediction the physicist bets in favor of is confirmed, we feel "strengthened in our conviction that the relations established by our reason among abstract notions truly correspond to relations among things" (1906/1954, 28) He also says of the reason and commonsense responsible for

AfU and TNC revealing the '*truths*' about theoretical classification progressively reflecting a natural classification, "…the truths which this common sense reveals are so clear and so certain that we cannot either mistake them or cast doubt on them..." (1906/1954, 104) But as I shall argue, this realist attitude did not amount to making unqualified realist claims; it gained legitimacy for Duhem through its role in justifying the pursuits of logical unity and novel predictions.

To be clear, Duhem did not, in Fine's words on Einstein, "espouse a global doctrine or a specific set of beliefs about reality". But he does seem to have had an attitude that favored the idea that the classification of laws by theory approached a real underlying order of nature. This realist attitude is, of course, only towards the *classification of empirical laws* brought about by theory, and not towards any other parts of theory. And to reiterate, this attitude and idea were not inferred, or even intuited *based on empirical success*. They were based on other intuitive considerations, and as I shall argue below, helped make sense of the physicist's activities.

Importantly, drawing on Fine's views on Einstein, Darling says this attitude *motivates* the pursuit of physics for Duhem: "...for Duhem a realist attitude is a natural form of belief for us, which inspires the practice of science..." (2003, 1135) I take it that Darling follows Fine closely here and (justifiably) takes Duhem not to have been a realist in the sense of taking theories to approach some metaphysical reality but rather a realist in the sense of having a positive attitude towards TNC for the reason that it motivates the practice of physics. The quote that Darling mainly uses to support her interpretation of motivational realism is the following:

> **J**: the physicist is compelled to recognize that *it would be unreasonable to work for the progress of physical theory if this theory were not the increasingly better defined and more precise reflection of a metaphysics; the belief in an order transcending physics is the sole justification of physical theory* (quoted in Darling, 2003, 1134, emphasis as in original)

I think this interpretation takes a very nuanced stand and comes closest among the modern offerings to understanding Duhem's position correctly. It rightly focuses on what Duhem took to be the overall attitude of the physicist toward pursuing theory, rather than on the truth of theoretical claims. However in my view, there are important differences between Duhem and Fine's Einstein. Fine says about Einstein's drive for pursuing a realist science,

> The idea of pursuing science… is not suggested, of course, in terms of *cognitive appeal*, but rather in terms of what motivates, enlivens, and gives meaning to one's activities. (Fine, 2009, Kindle Locations 2016-2017, emphasis mine)

For Duhem on the other hand, although interestingly he too talks of an *aspiration* for theoretic unity – which sounds like a "drive" – both this aspiration and TNC are *intuitions* imposed on the minds of physicists, and as I argue below, cannot be denied on grounds of reason and common sense – and so clearly have great 'cognitive appeal'. Duhem says it's commonsensically impossible to let go of the AfU:

> [this] feeling surges within us with indomitable strength; whoever would see in this nothing more than a snare and a delusion cannot be reduced to silence by the principle of contradiction; but he would be excommunicated by common sense. (1906/1954, 104)

And of the "common sense" that imposes the AfU and TNC on the mind of the physicist as quoted earlier, "…the truths which this common sense reveals are so clear and so certain that we cannot either mistake them or cast doubt on them..." (1906/1954, 104) Further, he says TNC is "imposed as certain" (1906/1954, 335) on the mind of the physicist.

So clearly, Duhem's position is epistemically much stronger than a motivational realist position. Motivational realism seems to be (a sophisticated form of) wishful thinking, but Duhem's idea is more rationally grounded: AfU and TNC are rational intuitions in the sense that they are grounded in good reason and are not purely wishful/ psychological.

Further, on AfU and TNC transcending logic he quotes Pascal in the following:

> Reason has, therefore, no logical argument to stop a physical theory which would break the chains of logical rigor; but "nature supports reason when impotent and prevents it from talking nonsense even at that point." (1906/1954, 104)

Note that the above quote holds that denying AfU and TNC amounts to nonsense. Importantly, what's at stake for motivational realism seems to be an "enlivened" and psychologically motivated pursuit of science. For Duhem too, what's at stake is motivation, no doubt – as evidenced by M above. But it's much more than that: if we deny AfU and TNC, *reason* and *commonsense* are at stake.

Fine's Einstein talks of a realist attitude as a drive, a motivating force rooted in our psychology. Duhem says similar things about AfU and TNC. But in addition he stresses the intuitive force of both. Importantly, Fine does not use the word "intuition" anywhere in the discussion of motivational realism and that itself I believe marks a significant departure from Duhem: motivational realism is not a hard-do-deny intuition that imposes itself compellingly on the mind of the scientist. Darling acknowledges that AfU and TNC were intuitions for Duhem, in fact quite strongly so, but fails to see that this sets him apart from Fine's Einstein in an important way. Fundamentally, an intuition has knowledge or belief as its object. Duhem's physicist has the intuition that physical theory should be logically unified and that its classification of experimental laws approaches a natural classification. Contrastingly, Einstein's motivational realism according to Fine, was just a driving force, a motivation, that gave meaning and purpose to pursuing science. Although Duhem's intuitions and Fine's motivational realism have a commonality in that they are both in a sense pre-rational – they are not arrived at by rational deliberation – they are just different things. While intuitions grounded in reason can play a motivational role, as they in fact did for Duhem, they are not completely characterized by that role. Further, there is more to the TNC story than intuition: it had a special relation with the pursuit of theoretic unity for Duhem, which I explain below.

As discussed above, Duhem points out that physicists yearn for unity and coordination in theory: the different parts of a theory should all be logically coordinated. If one confined herself to the methods of physics, then one would not demand of theory any logical unity, but most physicists work toward logical unity. Here we have a question similar to the one raised above about confidence in novel prediction. What justifies the pursuit of unity? For Duhem, it was again TNC. The following discussion in Duhem is in my view central to understanding the relation between TNC and the pursuit of unity. Duhem points out that the physicist does actively work to construct logically unified theories and that "history shows him that this aspiration is as old as science itself and that successive physical systems have realized this desire more and more fully from day to day". But as with the case of novel prediction above he says,

> …the study of the procedures by means of which physical theory makes progress does not disclose to him the entire rationale of this evolution. The tendencies directing the development of physical theory are not, therefore, completely intelligible to the physicist if he wishes to be nothing but a physicist. (1906/1954, 296)

And,

> **U**: If, on the other hand, he yields to the nature of the human mind, which is repugnant to the extreme demands of positivism, he will want to know the reason for, or explanation of, what carries him along; he will break through the wall at which the procedures of physics stop, helpless, and he will make an affirmation which these procedures do not justify; he will be metaphysical.
>
> What is this metaphysical affirmation that the physicist will make, despite the nearly forced restraint imposed on the method he customarily uses? He will affirm that underneath the observable data, the only data accessible to

> his methods of study, are hidden realities whose essence cannot be grasped by these same methods, and that these realities are arranged in a certain order which physical science cannot directly contemplate. But he will note that physical theory through its successive advances tends to arrange experimental laws in an order more and more analogous to the transcendent order according to which realities are classified, that as a result physical theory advances gradually toward its limiting form, namely, that of providing a *natural classification*, and finally that logical unity is a characteristic without which physical theory cannot claim this rank of a natural classification. (1906/1954, 296-297)

As before, I contend that here too, Duhem is in the business of making sense of and justifying the physicist's activity – here, the activity being constructing physical theory that has the feature of logical unity. As before, Duhem is saying that asserting that the classification of laws by a unified theory is approaching a natural classification, is what rationalizes the pursuit of theoretic unity – just as it rationalizes the physicist's confidently engaging in theoretical prediction of new experimental laws. (It might be worth noting that for Duhem these are not really two separate activities. The core activity at hand is classifying experimental laws by a logically unified theory, in the process of which the physicist encounters "empty drawers" to be occupied by new, hitherto unknown laws. So prediction of new laws is also just classification – it involves anticipatory fitting of hitherto unknown laws into existing classification schemes. This is supported by quotes like P.) A perfectly unified theory yields a natural classification, and so the more unified a theory is, the closer it gets to yielding a natural classification – so this perfect unity, as natural classification, is an ideal goal that theory tends towards.

Since justification has been a central idea, it would serve us well to take a closer look at it. First, Duhem is concerned with justification not in the traditional epistemological sense, of knowledge claims, but of the activity of constructing unified theories and making empirical predictions with conviction. Recall that J urges that "the belief in an order transcending physics" is the sole justification of pursuing[23] physical theory.

In several passages, Duhem talks about justifying the pursuit of (a unified) theory and of making predictions with it. Recall C (p.168) for instance. Further,

> The physicist, then, finds himself in an irresistable aspiration toward a physical theory which would represent all experimental laws by means of a system with perfect logical unity; and when he asks of an exact analysis of experimental method and what the role of physical theory is, he does not find anything in it to *justify* this aspiration. (1906/1954, 296, emphasis mine)

And:

> The physicist is then led to exceed the powers conferred on him by the logical analysis of experimental science and to *justify*[24] the tendency of theory toward logical unity by the following assertion: The ideal form of physical theory is a natural classification of experimental laws. (1906/1954, 297, emphasis mine)

---

[23] The quote, as some others, simply says 'justification of physical theory". But given everything else that Duhem says about the physicist's reasoning – for instance "justify this aspiration" in the first block quote above and "justify this confidence" back in C on p.9 – I think it is fair to interpret "justification of physical theory" as "justification of pursuing physical theory" – in fact I see no other way of interpreting the phrase.
[24] This supports my point above: theory by itself does not tend toward anything – obviously, Duhem is using 'justifying the tendency of a theory' as a shorthand for 'justifying the physicist's giving theory that tendency'.

What does it mean to justify an activity? With respect to pursuing unity in theory, as in U above Duhem seems to mean having an explanation/ good reasons to pursue the activity. What count as good reasons then? There might be various responses to that. In the present context of Duhem, here are four possible candidates for justification of the activity of pursuing unified theories, with some Duhem-esque responses:

1. The strong intuitive aspiration for unity.

   Response: An intuition can be a driving force, but does not itself justify action based on it.

2. History tells us that unified theories *work*, and have always worked.

   Response: This backs up our pursuit of unified theory, but is not a good enough justification.

3. Unified theories are aesthetically appealing.

   Response: This is not a good enough justification.

   These last two responses are supported by the following:

> …the order in which theory arranges the results of observation does not find its adequate and complete justification in its practical or aesthetic characteristics; we surmise, in addition, that it is or tends to be a *natural classification…* (1906/1954, 335, emphasis as in original)

Rather than 1-3 Duhem I believe takes the only convincing justification then to be:

4. TNC.

Similarly, as argued earlier the only justification for confidently making novel predictions is TNC. It looks to me as if Duhem has in mind a teleological justification: good reasons for pursuing physical theory consisting in an ultimate purpose. Duhem seems to be saying that the ultimate purpose of pursuing unity is natural classification and that this purpose justifies pursuing theory, i.e., makes it worthwhile. If we didn't think that the classification provided by a theory as logically unified as possible was approaching a natural classification, then it makes no sense to pursue unity, pursuing unity wouldn't have justification, in the sense that there would be no reason to pursue it. Similarly, if we didn't think theory was ultimately approaching a natural classification then it would be ridiculous of us to bet in favor of its predictions. This teleological aspect is what probably makes Duhem's position on TNC seem motivational. Sure, an ultimate purpose of pursuing an activity could very well motivate the pursuit. And I'm not here denying that TNC as well as AfU played motivational roles for Duhem. But I'm arguing that there is more than motivation to Duhem's story. For Duhem, TNC and AfU have a far greater cognitive and epistemic appeal than a simple, and perhaps wishful, motivation: they are imposed on the mind of the physicist by reason and commonsense. Further, TNC is really indispensible in a way that motivational realism is not: without TNC, the activity of pursuing logically unified theories would not make rational sense to the physicist.

There is an apparent inconsistency in Duhem that needs to be addressed. I said earlier that for Duhem TNC was an intuition. But I have also just shown that for Duhem asserting TNC is what justifies the physicist's pursuit of unity and confident novel predictions. Does the latter mean that the physicist comes up with TNC as a justification by rational deliberation? If so, there seems to be a tension: how can TNC both be an intuition, as well as a rational justification? I contend that there is no tension. In a nutshell, the point is that TNC is an intuition and is in that sense pre-rational, but the physicist's strong conviction in it is a result of post-hoc (pragmatic) reasoning.

To elaborate: there is a question underlying the whole discussion on novel prediction and pursuit of unity: What if we ignored or worked on giving up the TNC intuition? As in the quotes above the physicist who does not want to step outside of the methods her subject affords can do this – she has no reason to entertain the TNC intuition or take it to be reasonable or even important. But eschewing the TNC intuition leaves the physicist with no recourse to find any reason to pursue physical theory. Sure, it is driven by a strong intuitive aspiration for unity and an innate confidence in novel predictions, and both of these are also backed by the history of physics, but what makes these reasonable pursuits? What explains and justifies the rationale behind physicists' pursuit of these through history? The intuition for Duhem was a 'feeling' or 'apprehension', which are not guaranteed to be right. Should we fight it or disregard it? Why should we uphold it? The answer for Duhem was that something important is at stake: justifying key components of the very pursuit of physical theory. It is due to this that the physicist is compelled to embrace and affirm TNC. Passage U above attests to this.

Hence, while TNC is a spontaneous intuition, the physicist, in trying to deliberately reason out the rationale behind confident empirical predictions and pursuit of unity in theory and hence justify her engaging in both, *recognizes* or *realizes* that the TNC intuition helps in this regard, and hence affirms it. TNC is not actively proposed as a post-hoc justification for confidence in empirical prediction or pursuit of unity, but the physicist in trying to find answers to questions of rationale behind these activities – questions that physics does not help answering – realizes the cognitive force of TNC. The TNC intuition is firmly etched in the mind of the physicist, and when the physicist realizes this and affirms TNC, she is able to make sense of and justify her activities.

6.3.6 Hasok Chang and rationalizing epistemic activities

I now want to introduce a view of pragmatic rationality that I think helps us think about Duhem better. I believe the best way to understand Duhem's view is to situate him in a recent work of Hasok Chang (2009). As in Chapter 4, Chang advances a pragmatic, conditional view of rationality where belief in an "ontological principle" is necessitated by commitment to engaging in a related activity. Ontological principles don't purport to say anything objectively true of the world, but are taken to hold by us only insofar as we are committed to intelligibly engaging in relevant activities. E.g., counting requires the principle of discreteness: if we want to count things, we have to assume that the things being counted are discrete, for otherwise our activity will be unintelligible. Similarly, engaging in contrastive explanation requires that we don't deny the 'principle of sufficient reason': the principle that when there is an observed change, there is a reason behind it. It may seem that these principles are some kind of a priori eternal truths, but in fact, it is if and only if we have to intelligibly engage in the activity that the principle in question has to be taken to hold. If we deny the principle of discreteness while at the same time attempting to count, then our engaging in the activity becomes unintelligible. Outside of the activity, the principle has no call on us. Importantly for Chang, these principles are not objectively true per se: their necessity is entirely pragmatically grounded. Denying them would render engaging in the corresponding activity unintelligible. So if one is committed to an activity, one had better not deny a principle that is constitutive of the intelligibility of the activity. Otherwise, rationality – here intelligibility – is at stake.

Of course, when he says a principle need not be taken to hold outside the relevant activity, Chang does not mean something as trivial as that the necessity of ontological principles operates via some on-off switch: turning on when we engage in the relevant activities and turning off when we don't – that would be absurd. What Chang means, I think, is that if we believe in engaging in an activity in particular situations – i.e. if we think engaging in it is sensible, then we cannot deny the corresponding ontological principle irrespective of whether we actually engage in the activity – hence the word "commitment". Commitment to an activity need not mean actual engagement in it. I think the right sense of commitment here is commitment towards the meaningfulness or acceptability of an activity. If we're not committed in this sense, we need not be tied to the principle. If for instance, we don't believe in engaging in rational prediction in a particular situation for we don't find it acceptable – say we resort to divination, then the principle of uniform consequence, the principle that the same initial circumstance always has the same outcome – is not necessary for us. Since some activities like counting certain kinds of things (like pebbles and tables and chairs) are so basic and pervasive, the corresponding principles *seem* to be unconditionally necessary: we are *always* committed to the activity of counting say pebbles on a beach, in that we always believe that counting them makes sense (that it may be irrelevant or not of interest in some situations is beside the point), so we will *always* find the principle that pebbles are discrete, necessary, even when not actively counting them.

We should not let such cases mislead us. Chang's conditionality still very much holds: if (and only if) we are committed to an activity, the relevant ontological principle(s) is necessary for us. It is just that in some cases the antecedent of the conditional is always true, so the consequent is also always true: I suppose there are many things like pebbles whose countability we are always committed to, unlike other things like say water – with commitment redefined as above. So if the activity we are committed to, here counting pebbles, is to be intelligible, then the principle that pebbles are discrete is necessary for us. If counting pebbles was something we didn't always find acceptable/ meaningful to engage in, or further still if we had never come up with the idea of counting[25], we would have never found the principle of discreteness necessary.

Finally, as in Chapter 4, I contend that while intelligibility of activities is indeed an important goal, there are other goals we might have as well: not only do we want our activities to be intelligible, but we might for instance want them to be practically beneficial, or consistent with (some of) our other activities, or have some ultimate

---

[25] May be like C.I. Lewis' (1929, 252) jellyfish world that lacks arithmetic

meaning or purpose for us. For each of these goals too, we might have corresponding ontological principles paired with respective activities. Just like in Chang's scheme intelligibility is about a proper match between an activity and a principle, I think that any of these goals – practical benefit, being purposeful etc. – can also similarly be about a proper match between activity and principle.

Duhem's position seems to me to fit this picture quite well. Both Chang and Duhem are interested in the rationality of our pursuit of specific activities. For Duhem, the activity in question is physical theorizing – which includes constructing logically unified theories yielding classifications of experimental laws, and predicting new laws based on them. The principle in question for Duhem is TNC. For both Chang and Duhem, the principle/ thesis doing the rationalizing is not meant to entail any beliefs about the way the world really is. Duhem does not say that metaphysical reality *is* 'really' carved up the way we classify of experimental laws – he says rather that the physicist is compelled to make a metaphysical assertion – TNC – in order to justify her activities of constructing logically unified theories and confidently making novel predictions. For both Chang and Duhem the principles only help us make sense of our epistemic activities. (It is worth noting here that Darling points out that motivational realism shares this feature too – it denies that science approaches any truth(s) about the world, but rather holds that the belief that it does, motivates our practices.)

Since the question of where Duhem falls within the spectrum of realist positions has been a burning one, it might be worth saying a few more words about it. I have shown that for Duhem affirming TNC was indispensible for making sense of and justifying pursuing logical unity and confident novel prediction. Beyond that, we could keep arguing about whether or not he made TNC as a full-blown realist claim. I am inclined to say that Duhem did not have an unqualified, unconditional ontological commitment to TNC: although he explicitly calls it 'knowledge' and 'truth' (1906/1954, 334), these epistemic claims stem from the physicist's pragmatic rationality considerations and were moreover rooted in, to use some of his own words, 'surmise', 'apprehension', 'feeling' and 'intuition' (although these were rational for him). In other words, he makes positive epistemic claims which seem to reflect a realist attitude, but these claims are not epistemologically grounded in the traditional sense.

The strongest realist claim we can make here I believe is a Changian one: Chang holds that when we are successful at the activity we pursue, the ontological principles we assumed in carrying out the activity are indirectly vindicated (in some vague way) – for him that is as realist as we can get about the ontological principles. Similarly Duhem's view is that our conviction in TNC is strengthened when a novel prediction turns out to be right and when the logical coordination of theory becomes better and better. Duhem does sound more passionate about his realist attitude to TNC than does Chang towards his ontological principles, but I see both their rationales as quite similar. I am not too opposed to a realist reading of Duhem though and am happy to concede that he had a realist position, as long as we acknowledge his (pragmatically rational) basis for it.

How can one explain the *intuitive* status of TNC if it is contingent on the activity of pursuing unity? Intuitions are simply there (in the mind) – they are not assumptions we actively make in response to something we engage in. Similar to the discussion of Chang above, we can say that the TNC intuition does not turn on and off – the physicist does not propose/ believe in it at will, for according to Duhem, the physicist is *always* committed to constructing not only theories that are empirically adequate, but that classify empirical laws and are logically unified. The mind of the physicist[26] for Duhem is such that there has always been the drive to produce unified theories and confidently make novel predictions, so the TNC intuition is always there. Duhem himself says that the AfU and TNC intuitions are "inseparable companions" (1906/1954, 103) (as are activity and principle in Chang's scheme). This reinforces the point I'm making here – as long as there is AfU, there is TNC. Similar to my argument about Chang above, even if the physicist is not at a given time engaging in the pursuit of a unified theory, she would still not deny TNC for she is always committed to the unification in that she always thinks it is meaningful and makes sense to engage in it. So in a nutshell again, for Duhem, physicists' activity of pursuing logically unified physical theories and confidently engaging in novel prediction are rationally unjustified without assenting to TNC. But if the physicist were not at all committed to pursuing unity or engaging confidently in novel prediction – as is the physicist who does not want to step out of the confines of her subject as Duhem discusses – then there is no (pragmatically) rational compulsion to affirm TNC. (In fact, extrapolating, I would add that if there were not even the *intuition*

[26] See Chapter IV in A&S for a detailed discussion of Duhem's ideas on various kinds of minds and their characteristics and tendencies.

of pursuing unity and no urge at all to bet in favor of a novel prediction then the TNC intuition would simply not exist in the mind of the physicist – and this is consistent with Duhem's discussion of different kinds of minds and intuitions. This again lines up with AfU and TNC being companions – if there's no AfU, there's no TNC.)

But there is a key difference. I said above that intelligibility is just one of several goals we may have in carrying out our activities: while Chang is interested in the intelligibility of activities, Duhem is interested in justification*,* in the sense of finding a purpose for our activities. For Chang, we wouldn't be pragmatically rational if we tried to engage in unintelligible activities. For Duhem, we wouldn't be pragmatically rational/reasonable if we engaged in unjustified, purposeless activities. But both hold that in order to uphold this pragmatic rationality, we have to affirm certain principles/ theses for otherwise intelligibility and justification are at stake, respectively.

To summarize, according to Duhem, physicists pursue theories that: 1. Classify experimental laws by means of mathematical equations, 2. Anticipate new laws, 3. Are logically coordinated and not one incoherent heap. Duhem says this is what physicists do when they pursue physical theory, and what justifies all of this is asserting TNC. Further, for one who does not wish to engage in the activity of pursuing physical theory – one who wishes to remain indifferent to the outcome of a prediction, or the logical coordination of different theories, does not have to subscribe to TNC. For such a physicist, holding on to TNC would have no purpose. This again lines up well with

Chang's activity-based conditional rationality. But as an empirical point, Duhem says that no real physicist is like that.

In conclusion, *this* is how I read J – not as being indicative of a motivational realism.

6.3.7 Duhem's Natural Ontological Attitude

Arthur Fine famously argued for a position he called the Natural Ontological Attitude (NOA). Very briefly, the position is that we should have the same ontological attitude towards electrons and genes, as towards chairs and tables – that we treat scientific truths the same way we treat everyday truths. I won't go into the details of this position as it's not relevant here, but what's interesting is what Fine says *about* this attitude: he calls it an "attitude towards science we can live by". I think Duhem was also concerned with such an attitude: the attitude a practicing physicist could live by, an activity that would justify her activities. In fact, Ernan McMullin (1990) calls Duhem's view on TNC his 'natural ontological attitude', appropriating the term from Fine. To clarify, Fine's and Duhem's attitudes themselves are vastly different, but their view of the role of such an attitude – the meta-attitude if you will – seem similar.

What I also want to point out is an aspect of NOA that Fine draws attention to, that I think Duhem's views also share. Fine says,

> NOA may well make science seem fairly intelligible and even rational, but NOA could be quite the wrong view of science for all that. If we posit as a constraint on philosophizing about science that the scientific enterprise should come out in our philosophy as not too unintelligible or irrational, then, perhaps, we can say that NOA passes a minimal standard for a philosophy of science. (Fine, 2009, Kindle Locations 2423-2425)

In parallel, Duhem says,

> Everything therefore urges the physicist to postulate the following assertion: To the extent that physical theory makes progress, it becomes more and more similar to a natural classification which is its ideal end. Physical method is powerless to prove this assertion is warranted, but if it were not, the tendency which directs the development of physics would remain *incomprehensible*. Thus, in order to find the title to establish its legitimacy, physical theory has to demand it of metaphysics. (1906/1954, 298, emphasis mine)

We see that both Fine and Duhem are interested in the physicist making rational sense of her practices – making them intelligible/ comprehensible – rather than in a hardcore realist account of science where theoretical claims are taken to be (approximately) true. Like Fine, Duhem says that the physicist is led to assert TNC for it justifies her activities and makes comprehensible the larger communal practices – the development of physics at large – her activities are situated in. This is also interestingly reminiscent of Herbert Feigel's (1950) views about the impossibility of an ultimate justification (in the sense of validation) of logic: " for the vast majority of mankind logicality is primarily a means in the struggle for existence and in the pursuit of more satisfactory ways of living. These ends we pursue as a matter of stark fact; they are part of our human nature."

6.3.8. Conclusion

So what was Duhem's rationale behind TNC?

TNC is 1. Intuited based on the elegance and efficacy of theory, 2. Asserted and affirmed because it justifies physicists' expectation of successful novel predictions and the pursuit of logically unified theories, and 3. Vindicated by (predictive and unifying) success of theory.

Traditionally, philosopher-realists, particularly the NMA camps, have had an outsider view of science: they look at the results of science and ask what to make of it. On the other hand, Duhem is concerned with the physicist-philosopher. How can the physicist get on with her work everyday? What motivates and justifies her work, regardless of what the "correct" philosophical picture of science may be? The answer for Duhem was upholding TNC. Like for Chang upholding the principle of discreteness is a pragmatic rationality requirement for it makes intelligible the activity of counting, for Duhem upholding TNC is a pragmatic rationality requirement for it justifies the activity of pursuing physical theory. And like for Chang, the principle of discreteness isn't empirically grounded, for Duhem TNC isn't empirically grounded – it is metaphysical. But it does, I contend – à la Duhem – give the physicist *understanding* of her practice of theorizing by justifying the pursuit and giving it purpose.

Section 6.2 is being prepared for publication to appear as:

Bhakthavatsalam, Sindhuja. "Duhem, Natural Classification, and Structure"

A slightly modified version of section 6.3 appears as the following publication:

Bhakthavatsalam, Sindhuja. 'The Rationale Behind Pierre Duhem's Natural Classification' – Studies in History and Philosophy of Science (2015) 51 11-21

References

- Chang, Hasok (2009) 'Ontological Principles and the Intelligibility of Epistemic Activities' in Henk de Regt, Sabina Leonelli, and Kai Eigner, eds. *Scientific Understanding: Philosophical Perspectives* 64-82 Pittsburgh: University of Pittsburgh

- Darling, Karen Merikangas (2003) 'Motivational Realism: The Natural Classification for Pierre Duhem' *Philosophy of Science* 70 (5) 1125-1136

- Duhem, Pierre (1906) *The Aim and Structure of Physical Theory* Repr. 1954 Princeton, NJ: Princeton University Press

- Duhem, Pierre (1996) *Essays in the History and Philosophy of Science*, ed. and translation with Introduction by Roger Ariew and Peter Barker, Indianapolis: Hackett

- Feigl, Herbert (1950) 'De Principiis Non Disputandum ...' in Max Black, ed., *Philosophical Analysis* Ithaca: Cornell University Press Retrieved from http://www.ditext.com/feigl/pnd.html 05/26/2015

- Fine, Arthur (2009-02-25) *The Shaky Game* (Science and Its Conceptual Foundations S) University of Chicago Press. Kindle Edition.

- Gower, Barry (2000) 'Cassirer, Schlick and 'Structural' Realism: the Philosophy of the Exact Sciences in the Background to Early Logical Empiricism' *British Journal for the History of Philosophy* 8:1 71-106

- Ivanova, Milena (2010) 'Pierre Duhem's good sense as a guide to theory choice' *Studies in History and Philosophy of Science* 41 58–64

- Ladyman, James (2011) 'Structural realism versus standard scientific realism: the case of phlogiston and dephlogisticated air' *Synthese* 180 (2) 87 - 101

- McMullin, Ernan. (1990) 'Duhem's Middle Way' *Synthese* 83 (3) 421-430 Springer

- Lugg, Andrew (1990) 'Pierre Duhem's Conception of Natural Classification' *Synthese,* 83: 409-420.

- Lewis, C.I. (1929) *Mind and the World Order. Outline of a Theory of Knowledge* (Repr. 1956) New York, Dover.

- Psillos, Stathis (1999) *Scientific Realism: How Science Tracks Truth* Routledge

- Putnam, Hilary (1975) 'Mathematics, matter and method' *Philosophical Papers* Vol. 1. Cambridge: Cambridge University Press

- Worrall, John (1989) 'Structural realism: The best of both worlds?' *Dialectica* 43: 99–124.

- Worrall, John (2005) 'Miracles, Pessimism and Scientific Realism' LSE webpage 1–55. Retrieved from http://www2.lse.ac.uk/philosophy/WhosWho/staffhomepages/Publications/NMAandPIfinal.pdf 01/30/2014

- Worrall, John (2011) 'The No Miracles Intuition and the No Miracles Argument' in Dennis Dieks, Wenceslao Gonzalo, Thomas Ubel, Stephen Hartmann, Marcel Weber, eds. *Explanation, Prediction, and Confirmation* 11-21 Springer