

# In Defence of Constructive Empiricism

Abram Hertroys

Master's Thesis  
History and Philosophy of Science  
Graduate School of Natural Sciences  
Utrecht University

First Supervisor: Dr. F.A. Muller  
Second Supervisor: Prof. Dr. D.G.B.J. Dieks

November 20, 2013



*Feiten zijn de knippers waarmee onze  
in de culturele wind wapperende  
theoretische netwerken onlosmakelijk  
aan de waslijn van de wereld zelf  
zijn opgehangen.*

—André Klukhuhn (2008, p310)

# Contents

<b>Preface</b>	<b>v</b>
<b>Introduction</b>	<b>1</b>
Manifest versus Scientific . . . . .	1
Constructive Empiricism . . . . .	2
Outline . . . . .	4
<b>1 Scientific Rationality</b>	<b>6</b>
1.1 Introduction . . . . .	6
1.2 Stances in Stalemate . . . . .	7
1.2.1 There and Back Again . . . . .	7
1.2.2 Stance on Science . . . . .	9
1.2.3 Voluntarism and Stances . . . . .	11
1.3 Voluntarism Is Relativism . . . . .	13
1.3.1 Probabilistic Coherence . . . . .	13
1.3.2 Pragmatic Coherence . . . . .	15
1.3.3 Almost Anything Goes . . . . .	18
1.4 Reconceptualising Rationality . . . . .	20
1.4.1 Standard of Appraisal . . . . .	20
1.4.2 Evidential Effort . . . . .	21
1.4.3 Against Evidential Invulnerability . . . . .	24
1.5 The Scientific Context . . . . .	28
1.5.1 Empirical Evidence . . . . .	28
1.5.2 The Primacy of Experience . . . . .	32
1.5.3 Responsible Risks . . . . .	33
1.6 Against Manifestism . . . . .	35
1.6.1 Responsible Risks Revisited . . . . .	35
1.6.2 Making Sense of Science . . . . .	36
1.6.3 Judging Manifestability . . . . .	37
1.7 Conclusion . . . . .	39

<b>2</b>	<b>Miraculous Abductions</b>	<b>41</b>
2.1	Introduction . . . . .	41
2.2	No Miracles . . . . .	42
2.2.1	Putnam's Intuition . . . . .	42
2.2.2	Laudan's List . . . . .	43
2.2.3	Stance-Off . . . . .	45
2.2.4	Structure of the Inference . . . . .	49
2.3	Refining Abduction . . . . .	51
2.3.1	Explanans: Particularise . . . . .	51
2.3.2	Explanandum: Success . . . . .	52
2.3.3	Conclusion: Likelihood . . . . .	53
2.4	Loveliness and Likelihood . . . . .	56
2.4.1	Symptom of Truth . . . . .	56
2.4.2	The Bayesian Framework . . . . .	59
2.4.3	Objectifying Bayes . . . . .	60
2.5	Voltaire's Objection . . . . .	63
2.5.1	Explanation as Information . . . . .	63
2.5.2	Internal Resistance . . . . .	64
2.5.3	Context of Construction . . . . .	67
2.6	Conclusion . . . . .	68
<b>3</b>	<b>The Doxastic Policy</b>	<b>71</b>
3.1	Introduction . . . . .	71
3.2	A Puzzle of Actuality . . . . .	72
3.2.1	Two Nuclear Power Plants . . . . .	72
3.2.2	Judging Actuality . . . . .	73
3.2.3	The Letter in the Drawer . . . . .	74
3.3	On What is Actual . . . . .	76
3.3.1	Against Type-Actuality . . . . .	76
3.3.2	Against Object-Actuality . . . . .	78
3.4	Easing the Doxastic Policy . . . . .	79
3.4.1	The Revised Policy . . . . .	79
3.4.2	Counterfactual Observations . . . . .	80
3.4.3	Incorporating Non-Actual Observables . . . . .	83
3.5	A Humean Dissolution . . . . .	85
3.5.1	Committing to Counterfactuals . . . . .	85
3.5.2	Judging Actual Observability . . . . .	86
3.5.3	Janus-Faced Commitment . . . . .	88
3.6	Conclusion . . . . .	92
	<b>Conclusion</b>	<b>95</b>

# Preface

Before I studied philosophy, I was very impressed by sentences starting with ‘scientific study shows...’ and thought of science as a mysterious activity. I still am and I still do. Yet, in a different way, perhaps. After taking notice of Hanson, Kuhn and Feyerabend, I was so excited that I almost ended up in social constructivist quarters. Still, something bothered me. Nietzsche reminds us that the bee builds with the wax that he gathers from nature, but Bacon understood that man is not unlike the bee: we are equally stuck with nature’s material. Why else would the moon-sick Pierrot keep bumping his head?

The view of science that is the subject of this essay has become somewhat of a thematic thread through my programme since I picked it up in my second year. It was a pleasant discovery: I remember my annoyance by not comprehending the mystifying German nineteenth-century metaphysicians, the relief when I learned of logical positivism’s fierce criticism, and the disappointment in finding out that they, in turn, had become the piñata of today’s views. With the introduction of constructive empiricism, a viable alternative entered the arena.

In epistemology, I never got the urge to go external, certainly not regards justification; truth is already out of our hands, so if justification is rendered equally unrecognizable, we are left with only our beliefs and no guidance. Indeed, Worrall’s remark that ‘externalist epistemology’ is an oxymoron rings true to me. During the course, filled with Gettier-games, I read this liberating article on cutting the ties between knowledge and justification. The idea stayed with me, but I forgot the whom and whence. While writing this thesis, I finally rediscovered it (Foley’s “Skepticism and Rationality”) and, sure enough, it turned out to be inspired by Van Fraassen’s call for non-defensive epistemology.

This essay builds, fortunately, on much understanding I accumulated during my studies, but it seems appropriate to point out that especially Sections 1.6 and 3.4.1 draw upon my bachelor’s thesis, while the second Chapter improves on some ideas I put forward in an essay review of Lipton’s *Inference to the Best Explanation* and a paper on structural realism.

Two of the above papers were supervised by Dr. Janneke van Lith and I thank her wholeheartedly for her excellent and inspiring support. I also

cordially thank Dr. Jan Sprenger for introducing me to the world of probabilities. This thesis's greatest intellectual debt is to my supervisor Dr. F.A. Muller, whose first name will remain undisclosed. His comments—always meticulous and sometimes very amusing—were of great benefit to both content and spirit.

On a more personal note, I thank my parents for being there for me and generously understanding that I was sometimes unable to return the favour, absorbed as I was with the taming of this thesis. Thanks to my friends Jacinta, for her moral support, and Roald and Pjotr for helping me prepare the document for print. Lastly, I wish to thank the two girls that cheered up my life during the first few months of writing. Sarah, when you came to live with me, I initially abduced your existence, but increased the credence when I spotted you in the sink and in the oven. I do not know where you went, but I hope you are well. Thanks for not eating my cables. Надежда, ever hopeful creature behind the door one step from mine, before you moved on, you tried your best in preventing this thesis from coming to actuality. Helaasje, it is finished now.

# Introduction

## Manifest versus Scientific

Opening his 1927 Gifford Lectures at the University of Edinburgh, Sir Arthur Stanley Eddington told his students that he wrote his material while seated at two tables. He had not one but two tables, although the tables coincided. Table one is what we could call the *manifest* table. It is the table that everybody knows from their experience; that big, heavy thing made of solid oak, which we touch and push and paint. The other table is the *scientific* table. It is the same object, but it is comprised of mainly empty space and many little things in motion that we do not touch or push or paint individually. The elements are colorless, odorless and tasteless, and they cannot be observed with the naked eye. Science describes a whole world beyond the commonsensical world and this scientific image is not even similar to its manifest sibling. As Eddington had it, the manifest world is, in one sense, prior to the deep world beyond. The manifest world is the world with which we are familiar; the other is relatively alien.

The year after Eddington published his lectures (1928), a group of revolutionary scientists and philosophers printed a pamphlet titled “The Scientific Conception of the World” (1929). The manifesto was signed by the *Wiener Kreis*, which arose from the Ernst Mach society, chaired by Moritz Schlick. In the manifesto, Mach is credited with investigating how scientific claims can be reduced to, what may be called, their ‘manifest elements’, i.e. Russellian sense data, directly springing from what we can touch, push and paint. Scientific claims about the table’s elements need to be spelled out, at bottom, in terms of statements about the manifest table if they are to *make sense*. Insofar scientific claims cannot be analyzed in their *logical* and *sensible* (i.e. the empirically ‘given’) constituents, they are not scientific claims at all, but they are *senseless* and should be thrown at Hume’s anti-metaphysical bonfire.

What makes a sentence meaningful? Debates over the amount of angels that can dance on the tip of a needle are not connected to our sensory experience and are, hence, more akin to poetic expression than to science; the claims involved are dressed in the garbs of descriptive talk, but they are really empty and signify nothing. Our (non-tautological) knowledge starts and

ends with experience. The *verification-criterion of meaning* associated with logical empiricism holds that all and only claims that can be ‘in principle’ verified by sensory experience are meaningful. Yet, this criterion has fallen on hard times. For one thing, verifiability is *too strong* a requirement: for instance, *universal* claims cannot be verified by *particular* observations. For another, it is unclear how *dispositional* terms, like solubility, can be broken up exhaustively in terms of verifiable statements.

By the time that Quine published his “Two Dogmas of Empiricism” (1951), logical empiricism was already past its high point. The logical empiricist needs a distinction between *analytic* truths, propositions that are true solely in virtue of their *meaning*, and *synthetic* truths, which are propositions about matters of *fact*, in order to analyze scientific statements into their (tauto)logical and factual content. Quine argues that this distinction cannot be drawn. Hence, the reduction of scientific statements to their empirical content is impossible and, by the same token (p38), the empiricist verification-criterion of meaning cannot aid the explication of ‘sameness of meaning’ and analyticity.

Ten years later, during two lectures at Pittsburgh, Wilfred Sellars (who had previously debunked ‘The Myth of the (empirically) Given’ (1956)) revisited Eddington’s images: manifest versus scientific. But the sands had shifted. His stereoscopic recommendations notwithstanding, Sellars gives ultimate precedence to the *scientific* image, which is characterized by its postulation of supra-empirical entities. (1962) As Kuhn’s *Scientific Revolutions* (1962) dashed all hopes for empirical foundations, by the mid-sixties and during the next decade, the logical empiricist take on science waned away.

## Constructive Empiricism

In 1980, Van Fraassen’s *The Scientific Image* saw the light of day: a monograph in search of a revised and viable way to conceive of science in an empiricist spirit. One of the most important improvements is that he does not make a distinction between observational and theoretical predicates (those predicates that refer to sensory experience and those that are in need of analysis). Instead, the difference between what we can and cannot experience (by means of our unaided senses) is a difference in the world in relation to us, human observers: the difference between the *observable* and *unobservable* entities (objects, events, processes).

Our ways of talking about the world are thoroughly *theory-infected* and the project of reducing scientific theories to ‘pure’ observation statements was bound to fail. Rather, we should take our theories *literally*: if they involve claims about unobservable entities, these are not shorthands for observational talk, but literally meaningful and made true or made false by the

world. Yet, some of the models of a theory *represent* the phenomena: the actual observable entities. We can single out these *empirical substructures* in virtue of what they represent. Although science is theory-infected, we can restrict our belief in the theories to the belief that the empirical substructures account for the phenomena.

The distinction between the observable and the unobservable is *vague*, like portable and breakable, yet perfectly meaningful because there are clear cases and counterexamples: mountains are observable (not portable, arguably unbreakable), electrons and radiowaves are not. It is also an *anthropocentric* distinction: the ‘able’ in ‘observable’ signifies the human *ability* to observe, and what we, humans, can observe depends on our sense organs and brains. The distinction has no *ontological* implications: the existence of e.g. electrons does not depend on the human senses. However, it does have *epistemological* implications: our ‘will to believe’ (the sort and scope of beliefs we are prepared to adopt) can be informed by our beliefs about our *epistemic access* to the world. Since we are not directly in contact with unobservables, we can decide to remain neutral about their nature and existence.

Alongside the un/observable distinction, Van Fraassen differentiates between an *epistemic commitment* (belief) and a *pragmatic commitment*. The latter involves a certain engagement to—and immersion in—theories, which falls short of belief, but is characterized by the scientist deciding to use the theory, including its supra-empirical structures, for purposes of explanation and experiment design, and by his expectation that the research programme will remain fruitful in the future.

Van Fraassen proposes a novel, empiricist-minded view of science called Constructive Empiricism (CE). According to this view, science is aimed at the construction of theories that are adequate to the observable, i.e. *empirically adequate*, and full acceptance of a theory implies as belief only that the theory is empirically adequate. We are allowed to remain neutral about the supra-empirical implications of accepted theories, although acceptance also implies a pragmatic commitment. This view of the aim science and of the commitments involved in theory acceptance characterizes CE in a nutshell.

CE is embedded in an epistemological environment, consisting of two components: (i) an account of rationality and (ii) an engagement with the empiricist spirit. (i) The account of rationality, namely voluntarism, holds that any truly *coherent* philosophical position is rational and allows e.g. for inferences that go ‘beyond the evidence’. This implies that the scientist who *does* believe in the full truth of accepted science cannot be accused of irrationality, although his inferences (to *supra-empirical adequacy*) are not rationally compelling. (ii) The engagement with the *empiricist stance* centrally involves an admiration of the empirical sciences, including its emphasis on observation and experiment, and a disdain for metaphysical speculation. The adoption of the empiricist stance, thus, involves the decision to steer clear of abovementioned inferences to adequacy to the supra-empirical, and

accords well with the associated view of science: CE.

CE is contrasted with *scientific realism*. From the critical replies to *The Scientific Image* it became clear that realism has no agreed upon definition, but that the term covers a family of views that all involve the claim that science aims at something more ambitious than mere empirical adequacy. A naïve formulation of traditional realism is this: science aims at truth *simpliciter* and we have compelling reasons to believe that our best scientific theories, including their claims about the unobservable, are, indeed, approximately true. Realism, like CE, but *unlike* e.g. instrumentalism, takes scientific claims *literally*, but *unlike* CE, realism holds that inferences to the existence of unobservables are not only scientifically *licensed*, but part and parcel of science's *achievements*. The realist, therefore, attaches less epistemological significance to the un/observable distinction.

Consider the following. We have a series of (unaided) observations from which we infer a hypothesis about the unobservable entities involved. The hypothesis is not only able to unify and explain very well those observations, but also successfully predicts and explains a new series of observations. When this happens, we have, the realist might say, a good reason to believe tentatively that the hypothesis has something right about the postulated unobservables. How else is its success to be explained? And the present-day scientific theories that we find in the textbooks are the result of centuries of corrections and replacements with accumulating empirical and technological success, so we have even better reasons to believe in *their* approximate truth. How can the empiricist maintain that we should remain neutral about all but science's empirical consequences?

The debate between realism and empiricism, which was resurrected almost singlehandedly by Van Fraassen, is called *the realism debate*. This thesis is an attempt to defend CE by addressing some unsolved problems.

## Outline

I can see but two general ways in which a philosophical position—CE, voluntarism, realism—can be defended or criticized. From within, one can uncover hidden assumptions and tease out surprising conclusions to show that the position is *coherent* or not. If it is, the only resource left is to compare the assumptions and consequences with our pre-philosophical *intuitions*. That is the general method, here follows the plan of this thesis:

In Chapter 1, I sketch the outlines of the realism debate and argue that the arguments in support of CE rely on a prior commitment to empiricist principles: they are ineffective against the realist. If the discussion centers on the question whose portrayal of science makes better sense, it readily comes to a halt. Instead, the battle must move to the epistemological arena. Van Fraassen's liberal notion of rationality (voluntarism) renders both parties

rational and is not helpful in breaking the dialectical deadlock. The pivotal disagreement concerns the question of what is scientifically *permissible* evidence, and the arguments pro and contra the rival portrayals of science rely for their effectiveness on the assumed source of evidence.

With the global dialectical structure laid bare, some conceptual analysis is warranted in order to argue for an intuitively appealing account of rationality that is more demanding than voluntarism and, in particular, emphasizes responsible information-oriented activity by proposing two norms: *evidential effort* and *evidential vulnerability*. This account is applicable to the scientific context, or so I shall argue, by employing the appropriate notion of *scientific evidence*. If the only scientifically permissible evidence is experiential evidence, then the norm of evidential vulnerability at once renders realism epistemically *reckless*. Since views of science that suggest an aim that is more ambitious than inductive skepticism but less ambitious than CE are practically empty, CE is the proper view of science for an empiricist.

In Chapter 2, it is argued that a theory's *explanatory power*, the flagship epistemic virtue of realism, is no source of scientific evidence. More precisely: we have no more reasons to believe that explanatory loveliness is truth-conducive than to believe that, for instance, explanatory ugliness should guide our ampliative inferences. *Abduction*, i.e. the mode of inference that appeals to explanatory considerations, is utterly mysterious from an epistemic point of view. Rather, explanatory power should be regarded as a *pragmatic* virtue. Although supra-empirical sources of scientific evidence cannot be discarded *in toto*, the most prominent proposal is crossed out. In tandem with Chapter 1, this makes it plausible that CE is the best view of science.

Chapter 3 dismantles some attacks on the coherency of CE, both internal and in the light of general empiricism. I show that CE's doxastic policy, which equates full acceptance of a theory with the belief in its empirical adequacy, survives these charges unscathed. Much attention has been paid to the explication of observability, which looks like a concept that is unacceptable to the empiricist. I focus instead on the relatively neglected, other ingredient of empirical adequacy: *actuality*. The examination of this concept yields surprising puzzles, the solution of which illuminates both the importance of the pragmatic commitment to scientific theories as well as how the empiricist can practice what he preaches. Aside from being preferable by epistemological considerations, CE is consistent and coherent with the empiricist spirit. In sum, CE does not hold a candle for the empiricist spirit, but lights perpetual fireworks for it.

# Chapter 1

## Scientific Rationality

### 1.1 Introduction

Van Fraassen’s permissive account of rationality, called *voluntarism* (to be expounded below), places Constructive Empiricism (CE) and realism, if coherent, on an equal rational footing. The introduction of voluntarism, I claim, has put the realism debate in stalemate, *albeit* this is perhaps an inescapable consequence of the limited ‘human epistemic condition’. CE may be the superior view of science for someone with prior empiricist commitments, but for those who do not care for the empiricist spirit, realism could be the better option. The preferable view of science, thus, depends on one’s ‘epistemic outlook’ or the philosophical tradition wherein one feels at home. From within a tradition, when judging *ex cathedra* so to speak, one view of science may be the clear winner and the alternatives may be deemed inferior, but for the disinterested *voluntarist*—perhaps temporarily suspending his personal epistemic engagement—all coherent packages are rationally permitted.

Notes on terminology: with experience I mean *unaided sensory experience*, unless stated otherwise. ‘Evidence’ is used in a broad sense: as those beliefs that an agent finds epistemically relevant to some belief. In dialogue, evidence can be presented as a reason for belief. For instance, one might say, “I believe Mason is abducted by aliens, because I (believe I) witnessed it.” The agent entertains a belief about an event and offers evidence in the form of alleged experience. Aside from experience, there may be other ways of acquiring evidential beliefs, e.g. by intuition, clairvoyance or revelation. A different example: “I believe theory  $t$  instead of  $t'$ , because (I believe)  $t$  is the more unifying theory.” This statement can be interpreted in various ways. The speaker might motivate his theory choice without claiming that unifying power is epistemically relevant: he might give an explanation or excuse for his belief, i.e. his reason for belief is *pragmatic* instead of *epistemic*. Or he might claim that unifying power is, somehow, a symptom of truth. In the

latter case, the feature is presented as evidence: it is presented, not merely as an excuse, but as a *good* or *epistemic* reason for believing it.

Van Fraassen has coined the term *stance* to refer to certain philosophical traditions. A stance may involve beliefs, but more importantly includes values, commitments, a choice of concepts and other non-doxastic ingredients. He warns against the danger of a semantic inflation that also troubled Kuhn's 'paradigms', lets 'stance' signify a position, standing place, vantage point or posture, and chooses to dub his own stance 'the empirical stance' "to characterize the sort of epistemic policy that empiricists display as a paradigm of rational inquiry."<sup>1</sup> (2004, p171–8) I use the term *empiricist* stance. A practicing scientist is neutral or not concerning claims about the supra-empirical, whereas a philosopher endorses a view of science, such as CE. Since a stance is philosophical and encompasses the adoption of an *epistemic* policy, the term 'empiricist' is better suited. An 'epistemic policy', whether value-driven or motivated otherwise, leads to decisions on what sort of evidence one takes into account, i.e. to an *evidential* policy. This may be motivated by (empiricist) *reflections on* our epistemic access and behaviour, i.e. epistemology.

In this Chapter, we take two rides on a carousel centred on the pivotal point of disagreement in the realism debate. In the *first* run, I show that the strongest arguments in favour of CE presuppose a prior empiricist engagement: from a justificatory view point, CE and the empiricist stance can be taken as a *package deal*. This package may be *coherent*, but coherence allows for many rival packages, and these are all ratified by voluntarism. After scrutinizing voluntarism, I propose a sketch of an alternative account of rationality, which puts *evidence* and *responsibility* centre stage. The *second* ride shows that the new notion of rationality brings the parties closer together, but in the end, the dialectical stalemate cannot be solved. The remaining disagreement concerns the characterization of scientifically *permissible* evidence. On the bright side: for an empiricist, CE is the most responsible and most *sensible* view of science.

## 1.2 Stances in Stalemate

### 1.2.1 There and Back Again

Van Fraassen's thinking has made an interesting move: from a view of science (1980b)—which, by stating the criteria of successful science, *en route*, differentiates beliefs that are scientifically supererogatory from those which are in accordance with the aim of science—via a very general and permissive

---

<sup>1</sup>Immediately adding that "I may be chided then for naming the book as I did rather than 'The Empiricist Stance'. I am sure the reader can defend my actual choice here as well as I can." (2004, p178 fn13)

‘voluntarist’ account of rationality (1989), to the rationally optional empiricist stance (2002), which, in turn, should offer a hospitable, more general environment for the view of science.

A view of science needs an epistemological embedding: it needs to be backed up by reflections on our epistemic access and behaviour. Van Fraassen holds that the aim of science is empirical adequacy (EmpAd)—roughly: truth about the actual observable—and frames realism as stating that the aim is truth across the board. Now, in each other’s eyes, the opponent misunderstands the aim of science, or is revisionary, while their own view covers the right aim. An *interpretation* also functions as a *delineation*: anything that does not meet the proposed characteristics is ruled out as the intended subject of portrayal. To resolve the dispute between two interpretations, we are typically in need of a third one, against which to compare the rivals. Since an ‘objectively neutral’ description of science is a chimera, however, we need a notion of science that is either very widely endorsed and hardly deniable<sup>2</sup>, or at least agreed upon by both parties.

CE and realism are in broad agreement on what to characterize as ‘scientific activity’, but this activity seems unable to adjudicate between the rival views: both can make sense of it in their own way, without obvious omissions. (Historically, the criticism on CE was that it relegated much activity to the realm of scientific *pragmatics*, but this objection can be turned around, such that realism misinterprets e.g. experiment-oriented behaviour.) Both views are able to cover what is generally agreed upon as scientific activity, but they interpret the activity in rival ways: as aimed at a different goal. Hence, the views should be backed up by an answer to the question: why is yours the *proper* aim? Added thereto that realism typically includes the thesis that we have good reasons to believe that some suitable selection of scientific theories is approximately true about the unobservable, this begs for a relocation of the debate to the epistemological arena.

CE was initially backed up by the thesis that experience is our exclusive source of information about the world. (1985, p253, 258) Experience can offer information only about the actual observable (‘the empirical’), so we had better restrict our belief to claims about the empirical realm; any extra beliefs are evidentially invulnerable. That is, of course, the empiricist tenet. Yet, according to the voluntarist<sup>3</sup>, we are not rationally compelled to limit our beliefs thusly; we may freely believe as we like as long as we remain *coherent*. The extent of our underdetermined belief is a personal affair and

---

<sup>2</sup>Van Fraassen (1994, p190) appeals to a shared understanding of science. This is a sensible move: rival interpretations of ‘martial arts’ could quarrel over the proper classification of thumb wrestling (it satisfies the current Wikipedia-definition), but if some interpretation excludes *savate*, it is out of the competition.

<sup>3</sup>In my presentation of voluntarism throughout this Chapter, I have drawn from Van Fraassen (1980a; 1983a; 1984; 1985, p246–281; 1989, p129–214; 2000; 2002, p74–90; 2004; 2007, p338–355; 2011).

our courage to believe at the risk of being wrong is driven by decisions as to which beliefs we find valuable. The empiricist rejects realist reasons for belief in the extra-empirical, because he considers this courage a farce, but also thinks that this does not render realism irrational.

Recognizing that the empiricist thesis is circular if based on experience (Van Fraassen 1995a), and else self-defeating (Van Fraassen 1994a)—either devoid of information itself or implying extra-experiential information—the emphasis on experience and the dissatisfaction with metaphysics were transformed into a ‘stance’: a cluster of attitudes, commitments, values, beliefs and what not. This stance provides an implicit *epistemic policy*, akin to the empiricist thesis, but is allegedly not self-defeating because it is not strictly a thesis. In the light of the libertine account of rationality, the empiricist stance is, obviously, optional. Thus, CE is an option among various rational alternatives, but insofar as it was supported by an account of our epistemic access to the world, this support has been rendered equally optional. Bottom line: why should we believe that empirical adequacy is the proper aim of science?

### 1.2.2 Stance on Science

Let us reconsider two initial arguments for CE. The *first* argument is that CE is superior to realism in *making sense of science* and scientific activity, and it does so without inflationary metaphysics.<sup>4</sup> The rebellion against metaphysics is only explicable and endorsed from the stance, so this provides no argument for the package. The appeal to an intelligible portrayal of scientific activity is *prima facie* flat out circular, but it may be persuasive if it can be applied to a practice that both parties recognize as scientific; it has to proceed from *common ground*. As such, it could e.g. select a universally admired episode from the history of science and show that it evidently revolved around EmpAd instead of truth.

Now, I am not aware of any case studies presented by Van Fraassen to illuminate how CE makes *better* sense of science than realism—he does present numerous examples to show that it makes good sense *simpliciter*.<sup>5</sup>

<sup>4</sup>Van Fraassen (1980b, p73) and Monton and Van Fraassen (2003, p421). Also relevant to this argument are Rosen (1994) and Van Fraassen (1994b).

<sup>5</sup>Instead of saying that CE is superior *because* it involves no metaphysics—where this is broadly characterized as the speculation about ‘what goes beyond’ our experience and, thus, begs the question—CE might simply claim that it makes *equally good* sense of science with *less* epistemic risk than the realist. I think the realist should agree that the empiricist takes less risk, but he could retort that this attitude is *overly modest* in the light of *his* aim of science.

This still does not explain what Van Fraassen means with making *better sense*, although the following quote suggests it is indeed intended to not hinge on anti-metaphysical premises: ‘Even if some constructive empiricist were to embrace modal realism [...] she could still argue that constructive empiricism makes better sense of science than realism does. It is here—regarding how to best make sense of science—that one finds a central

(The comparative judgment cannot appeal to the assertion that CE steers clear of metaphysics; that selling point cannot be a *differentium* and simultaneously stand on common ground.) It is hard to see how the comparative evaluation might be established. Since the rival views of science are not intended as sociological descriptions, anecdotal evidence from actual practitioners is not convincing. Moreover, the realist may admit that experience has the final say, in the sense that conflicts between theory and observations are, in the long run, to be settled in favour of the latter, and maintain that other sources of evidence come in to break experiential ties. Also, CE does not deny that extra-empirical considerations play an important pragmatic role in scientific practice. The question which considerations count as *pragmatic* and which as properly *epistemic* is prior to the effective dissection of case studies.

The *second* argument in support of CE is, what Teller has dubbed, the *pointless epistemic risk* argument (as found in Van Fraassen 1980b, p72–3; 1985, p252–5). The realist belief in truth and the empiricist belief in EmpAd both go way beyond the evidence; they are ampliative epistemic commitments; undetermined by the experiential evidence—and when an EmpAd theory involves empirical generalizations, as it typically does, its EmpAd can never be conclusively established. The difference, however, is that a theory is only vulnerable to the extent that it is empirically vulnerable. Therefore, both epistemic commitments are equally vulnerable, but since the falsity of a theory can only be concluded by way of its empirical inadequacy, the realist belief is not independently vulnerable. Hence, the realist commitment has a greater risk at being *false*, but has no greater risk at being *falsified*; the realist risk is pointless.

The crucial claim in the above argument is, of course, that a theory is only vulnerable insofar it is empirically vulnerable. The opponent may simply retort that theories are vulnerable to additional evidence, e.g. as provided by comparing theoretical virtues and vices.

Teller (2001) and Cartwright (2007) consider ways of defending the primacy of experience by pointing out its special status: experience is imposed upon us, we are not able to deny it, and the empirical realm is “the only part which the world forces us to take seriously.” (Teller 2001, p128) These remarks seem to be independent of one’s stance and make it plausible that a certain emphasis on experiential evidence is to be recommended. However, they do not successfully advocate experiential *hegemony*. Note that experience is not always unequivocal and our beliefs about it are fallible, so this ‘imposing’ force needs to be qualified; there is some room for disagreement on experiences. The opponent can hold that certain extra-empirical beliefs are both equally uncertain and revisable, and with equal force imposed by

---

motivation, arguably the main motivation, for constructive empiricism.” (Monton and Van Fraassen 2003, p421) Regrettably, this statement is not clarified.

the world. It seems, then, that the pointless epistemic risk ‘argument’ is not a stance-independent argument, but rather an expression of the stance.

The above arguments in favour of CE are not effective against someone who rejects the empiricist stance. CE is not only backed up by a stance, but it seems that this stance is centrally a *stance on science*. It includes a cluster of commitments, attitudes and values, but *to be empiricist*, it should, somehow, give rise to the assertion that the empirical sciences are a paradigm of rational inquiry, which according to Van Fraassen, “expresses first of all a central, perhaps the main, attitude integral to empiricist epistemologies, and it trades implicitly on an assertion to the effect that the practice of the empirical sciences instantiates an epistemic policy to be admired and emulated.” (2004, p179) This admiration is meaningful only if ‘empirical’ and ‘science’ are; the realist also admires science, humbled by those empirical experiments which disclose the unobservable. The same applies to the propagated ‘disdain for metaphysics’: this disdain can only be directed if we have a prior understanding of ‘the empirical’.

The object of empiricist admiration is, obviously, to be construed in the way the empiricist conceives of it. Therefore, the stance is already wedded to a specific view of science: “In the case of an empiricist, this stance would, for instance, involve a characterization of what science is (in my opinion it is a pursuit of empirical adequacy) [...]” (Van Fraassen 1995a, p86) If CE is the main ingredient in what constitutes the stance, the epistemic outlook and the view of science might better be evaluated as a unity. On the other hand, above remarks may just abbreviate the claim that the empiricist stance involves certain components that would, if spelled out properly, lead to (an embrace of) CE.<sup>6</sup> Regardless, CE is for its viability parasitic on the stance, so from a justificatory point of view we might as well approach it as a *package deal*. Our question then becomes: why should we be persuaded to opt for this package?

### 1.2.3 Voluntarism and Stances

The introduction of voluntarism, Van Fraassen’s libertine account of rationality, only deepens the impasse. Voluntarism is put forward as an unavoidable analysis of the human epistemic condition. Its main pillar is the claim that *coherence* is the sole and sufficient criterion of rationality. Coherence concerns belief-webs, wherein these beliefs may come in degrees. The voluntarist notion of coherence is broadly construed: it stretches the probabilist

---

<sup>6</sup>But again, compare “Such a slogan or putative doctrine as ‘Experience is our sole source of information’ may still have some significance for the empiricist [...]. In my view, however, *it comes along as a belief with one’s commitment to empirical inquiry* [...]. That is, the commitment is not based on that belief; instead there is a certain kind of pragmatic incoherence in having such a commitment while denying or expressing disbelief or doubt with respect to that statement.” (Van Fraassen 2004, p173, my emphasis)

norm of synchronic coherence to a more encompassing norm of pragmatic coherence, based on the view of belief as future-directed engagement. (This conception of coherence is explored further below.) Coherence is, for a large web of beliefs, not a weak constraint, but since different coherent webs may wildly vary in their content, the account has been called very *permissive*.

This permissiveness is mainly motivated by the perceived failures of more demanding epistemologies. Coherence is a feature of belief-systems, so in traditional epistemological terms it is an *internalist* constraint—it takes the believer’s perspective as its starting point. Voluntarism denies foundational security. Thus, it gives up on Cartesian foundationalism, admits fallibility, and refuses to get bogged down by the sceptical challenges. It also rejects the less securing but still foundational efforts aimed at enriching the probabilist framework with objective prescriptions for recipes for induction. The attempts at spelling out an objective relation between evidence and theory in terms of warrant, justification and confirmation are deemed part of a bankrupted ‘defensive’ project.<sup>7</sup>

Instead of following William K. Clifford, who said that “[i]t is never lawful to stifle a doubt; for either it can be honestly answered by means of the inquiry already made, or else it proves that the inquiry was not complete,” (1879, p187) Van Fraassen expands on William James, who retorts that “[s]cience would be far less advanced than she is if the passionate desires of individuals to get their own faiths confirmed had been kept out of the game.” (1896, p340) According to James, the epistemic aim is two-fold: believe all truths, believe no falsehoods. To satisfy the first desideratum, we should believe every proposition, because the truths are among these; to satisfy the second, we had better believe nothing at all. The desiderata pull in opposite directions, and for lack of an objective theory-evidence-relation, the balancing out is a personal affair. The decision whether or not ‘to have faith’ (to believe) is driven by ‘the passionate desires’ (one’s values), and this decision cannot impugn on one’s rationality.<sup>8</sup>

Hence, the traditional justified, true beliefs are transformed into a matter of ‘technique, courage and luck.’ (Van Fraassen 2000, p272–9) Courage is displayed by believing something at the risk of believing a falsehood. Luck encodes that the world needs to cooperate for this fallible and uncertain belief to be true. Technique means no self-sabotage, which is explicated in terms of

---

<sup>7</sup>Van Fraassen has attempted to give a global argument against the probabilistic coherence of adopting ampliative algorithms, i.e. rules for induction (1989, p160–70), but this argument essentially depends on the alleged incoherence of one’s liability to hypothetically accept a so-called diachronic ‘Dutch Book strategy’, and since Van Fraassen has subsequently abjured such arguments (1995b, p9), I suspect this line of criticism is withdrawn. In the next Chapter, we further explore arguments concerning ampliation.

<sup>8</sup>Although voluntarism may be naively understood as the position that we have full, voluntary doxastic control, this is emphatically not how Van Fraassen intends his position. (1984, p236, fn3) Rather, it expresses the view that belief is something akin to an engagement.

‘pragmatic coherence’ (expounded in Section 1.3.2). Thereby, rationality is pulled apart from traditional analyses of knowledge, which keep on fiddling with the justification criterion in fruitless attempts to win the Gettier-game.<sup>9</sup> Voluntarism refuses to be bothered by these scenarios and concedes that we might have fully rational beliefs which are wide off the truth-mark if we happen to be unlucky. We have to work with what is in our hands; our values and technique.

Accordingly, epistemological stances are evaluated by their coherence; every truly coherent stance is rational. From within a stance, one is committed to upholding it (this is part of pragmatic coherence). Temporarily suspending this commitment, we recognize that all coherent stances are on an equal rational footing. Notably, Van Fraassen shifts back and forth between engagement and relativism in a single clause: “while I respect the rationality of those who prefer to have those supererogatory beliefs [. . .].” (2004, p168) The quote begins with typical voluntarist open-mindedness, but the switch is signalled by the disqualification ‘supererogatory’: unneeded in the light of CE—which is motivated by the stance with its implicit epistemic policy rendering such beliefs evidentially invulnerable and *therefore* supererogatory. CE hinges on the empiricist differentiation of pragmatic and epistemic considerations; if we had extra-empirical sources of evidence, these should surely be accounted for in our paradigm of rational inquiry. Voluntarism renders the empiricist stance rational, but permits us to adopt any other coherent epistemic outlook and ensuing view of science.

This libertine account of rationality makes it hard to see how the package deal can be sold to the unwilling. Taking a stance comes down to making something akin to Kierkegaard’s *radical choice*.<sup>10</sup> The portrayal of science depends on one’s notion of information or scientific evidence. Different stances give different answers, preach self-affirming arguments and fail to yield converts on rational grounds.

## 1.3 Voluntarism Is Relativism

### 1.3.1 Probabilistic Coherence

In this Section and the next, we explore the extent to which coherence restrains our doxastic households. Then, we show that the minimalist account leads to relativism.

Probabilistic coherence means conformity to the probability calculus. Although there is much debate about the proper interpretation of the calculus,

---

<sup>9</sup>For an insightful analysis of the overly defensive attitude prompted by scepticism and Gettier-scenarios, see Foley (1990; 2005). He gives a comprehensive characterisation of why and how to turn to non-defensive epistemology.

<sup>10</sup>This metaphor is also found in Rosen (1994, 156).

the axioms themselves, as given by Kolmogorov ([1933] 1956, p2), are relatively uncontroversial.<sup>11</sup> Let  $\Omega$  be a non-empty set of elementary events and  $\mathcal{F}$  a set of subsets of  $\Omega$  that contains  $\Omega$ ,  $\emptyset$ , and is closed under complementation and union.  $P$  is a function from  $\mathcal{F}$  to a real value in the interval  $[0, 1]$ .

1.  $P(A) \geq 0$ , for every  $A \in \mathcal{F}$ .
2.  $P(\Omega) = 1$ .
3.  $P(A \cup B) = P(A) + P(B)$  when  $A \cap B = \emptyset$ , for all  $A, B \in \mathcal{F}$ .

Let us call  $P$  a probability, which, thus, assigns a value from the interval  $[0, 1]$  to events  $A, B$ , etc.  $\in \mathcal{F}$ . The third axiom, *finite additivity*, demands that the probability of the union of disjoint events is the sum of their individual probabilities.

For example, if we model the possible disjoint outcomes of a coin toss as  $\Omega = \{\text{heads}, \text{tails}\}$ , we cannot assign, for instance,  $P(\text{heads}) = 0.7$  and  $P(\text{tails}) = 0.7$ , since this would violate either the second or the third axiom, i.e. it is incoherent.<sup>12</sup>

Kolmogorov used *events* to present the possibilities to which values are assigned, but probability can also be assigned to *propositions*, in which case tautologies get maximum probability 1 and a disjunction of incompatible propositions is the sum of their probabilities. (cf. Weisberg 2009, p478–9)

The *conditional probability*, the probability of  $A$  given  $B$ , is a new probability measure, obtained from  $P$  and an event  $A \in \mathcal{F}$ , as follows. (Kolmogorov [1933] 1956, p6):

$$P(A|B) = P(A \cap B)/P(B), \text{ when } P(B) > 0.$$

We throw a die with the possible outcomes 1 to 6 and we assign to the possibilities a uniform probability,  $1/6$ , so  $P(\Omega) = 1$ . The unconditional probability of throwing a 6 is  $1/6$ , but *given* that the outcome is an even number, the probability is  $1/3$ : the intersection of 6 and the even numbers is 6, which has probability  $1/6$  and the probability of the union of even outcomes is  $1/2$ .

$$P(6|\text{even}) = P(6|2 \vee 4 \vee 6) = \frac{P(6)}{P(2 \vee 4 \vee 6)} = \frac{1/6}{1/2} = \frac{1}{3}$$

<sup>11</sup>Two vexed questions are (i) whether we should include a fourth axiom on countable additivity, generalizing the third axiom to cover infinite sets (Williamson 1999), and (ii) whether we had not better, like Popper ([1934, 1959], 2009), take conditional probabilities as primitive and define unconditional probabilities in terms of them (Hájek 2003).

<sup>12</sup>If these values are translated to hypothetical bets, such that you would judge fair a bet that pays 1 unit (dollar, utility point, candy bar) and is bought for 0.7 on both tails and not-tails, a clever bookie takes both bets with you and you are guaranteed to lose 0.4 come what may. This is known as a Dutch book and liability to a Dutch book equals probabilistic incoherence. The classical sources for this pragmatic justification of the norm of probabilistic coherence are Ramsey ([1926] 1931) and De Finetti ([1937] 1992).

Voluntarism asserts that our rational graded beliefs (rational credences) should be probabilities, i.e. conform to the axioms, i.e. be coherent. Why suppose that our credences should be coherent on the penalty of irrationality? Van Fraassen (1983a) proposes a frequentist analogue to the pragmatic Dutch book justification of the normativity of probabilistic coherence, in terms of epistemic vindication: potential perfect fit with the actual relative proportions, i.e. *calibration*.<sup>13</sup> In a nutshell: since these proportions obey the probability calculus, our expectations had better as well, if we are not to sabotage potential calibration from the outset. This basic requirement allows for the possibility of rational disagreement in the light of equal evidence. Thus, Van Fraassen denies that our credences “will have been arrived at in a rational manner, exactly if the input (background beliefs and information) determines via the dictates of rational deliberation, a uniquely right answer”. (1983a, p299)

The values in the interval  $[0, 1]$  are a formalistic device to ‘capture’ or present our credences and reason with them. Yet, workaday conversation suggests that not any graded belief can be assigned a sharp value. We say, ‘sadly, it seems not unlikely that she committed adultery,’ not, ‘I give her 0.83’ (with a sharp infinite string of zeros appended). To mend this descriptive inadequacy, we may introduce *indeterminate credences*. There are various ways in which this can be made precise: the credence function assigns not values but intervals, credences are represented by sets of functions, or we divert to comparative credences, which may even be devoid of a determinate range of values.<sup>14</sup> Although such moves yield their own peculiar difficulties, we gloss over this and proceed under the convenient fiction that our credences are determinate.

### 1.3.2 Pragmatic Coherence

Probabilistic coherence, sometimes called *synchronic* coherence, is extended to include a temporal dimension. The norm of *diachronic* coherence is usually motivated by showing that if one’s credences, including future-directed credences, at a specific time do not conform to the norm, one is liable to accept or judge fair (hypothetically) a certain combination of bets throughout time which results in certain loss. This so-called diachronic *Dutch strategy* is famously presented by Teller (1973) and attributed to Lewis. This argument

---

<sup>13</sup>The relative frequency of a  $p$ -type outcome is the proportion of  $p$ -type outcomes relative to the total set of actual outcomes of trials of the reference class. The relative frequency of some coin (reference class) toss (trial) landing heads (outcome) is the proportion of head-outcomes in the set of all actual tosses of the coin (past, present, future, and perhaps including actual outcomes of ‘relevantly similar’ coins).

<sup>14</sup>Aside from the more psychological considerations, there are also mathematical arguments for indeterminacy. An overview is found in (Hájek and Smithson 2012, p34). For sets of functions, see (Levi 1974; Van Fraassen 1980a; Jeffrey 1983; Gaifman 1988), for comparative credence, see (Bartha 2004).

is typically used to motivate the Bayesian conditionalization rule: if you are committed to apply this rule to update your credences when new evidence comes in, you are immune to these losing betting strategies. Van Fraassen initially used a similar argument to establish his principle of reflection. Since then, he has abjured appeals to Dutch strategies (1995b, p9), and reflection is deemed part of his broader notion of *pragmatic* coherence, which may be summarized as the view that adopting a belief is similar to adopting an engagement.<sup>15</sup>

The term ‘voluntarism’ was, as far as I was able to determine, first used by Van Fraassen in his (1984), wherein he introduces his in/famous principle of reflection. Voluntarism is proposed as a ‘solution’ to ground or explain the alleged normativity of the *principle of reflection*. This principle is the demand that if one believes that one’s graded belief about some proposition equals  $x$  at some later time, it should equal  $x$  at present, on the penalty of vulnerability to a diachronic Dutch strategy. Voluntarism, then, explains this diachronic coherence requirement by interpreting the (internal) expression of a graded belief as some kind of engagement akin to a promise: beliefs should not be held merely momentarily but involve the intention to stick with them in the future. Of course, informational circumstances and beliefs change, but it is purportedly pragmatically incoherent to avow a belief and add that you will change your mind tomorrow. The term ‘voluntarism’ is supposed to highlight that beliefs are “a matter of cognitive commitment, intention, *engagement*. Belief is a matter of the will.” (*Ibid.*, p256)

The principle of reflection is meant to demonstrate, contra Bayesian orthodoxy, that we can freely embrace novel hypotheses and throw doubt on evidential input without thereby becoming incoherent; we may rationally change our beliefs, not bounded by conditionalization, and without becoming necessarily vulnerable to diachronic Dutch strategies as long as we obey

---

<sup>15</sup>Although Van Fraassen considers voluntarism and probabilism as separate components of his new epistemology (2004, p182, fn18), the two norms of coherence may be subtly intertwined. The calibration argument appeals to vindication, which, on an actual frequentist interpretation (1983a, p295), can only make sense when a series of (actual) trials is considered. A belief such as ‘it seems as likely as not to me that the coin will come up heads,’ cannot be vindicated by a single trial (Van Fraassen rejects the propensity account, according to which there are objective single-case chances). Calibration is measured against a series of outcomes relative to some reference class (*Ibid.*, p304–5). Yet, the scoring procedure presupposes that we keep our expectation of the proportion of some designated outcome relative to some reference class *constant* during the sequence that is used to measure calibration (cf. Van Fraassen 1989, p159; Hoefer 2012, p447). These proportions obey the probability calculus, so our expectations had better as well. This argument can only trickle down to single-case expectations if these are stable for some  $>1$ -run. Hence, the parallel to promises. (Van Fraassen 1983a, p296–7) The frequentist justification of synchronic coherence seems arrived at via the detour of diachronically stable, committed beliefs—incidentally, the second, broader kind of coherence alluded to in connection to the principle of reflection. For further complications regarding a single-case norm of rationality based on a set of outcomes, see Hájek (unpublished).

reflection. Initially, voluntarism is supposed to be a solution to a puzzle created by a diachronic Dutch strategy: the puzzle suggests the *prima facie* untenable reflection principle, which can only be explained by interpreting belief as engagement (1984, p236). But since Van Fraassen has meanwhile forsworn Dutch strategies, voluntarism has to carry the full weight of the counterintuitive reflection principle.

This counterintuitiveness is illustrated by Christensen (1991), who considers scenarios in which it is clearly not rational to conform currently to one's expected expectations: e.g. when you believe that you will be intoxicated tomorrow. From the supposition that my future intoxicated expectation that I can fly is high, *it should not follow* that it should be high at the sober present. Indeed, *that* engagement would be irrational. Yet, the principle of reflection should have a principled, non-circular way of excluding such—and less radical—counterexamples.

One suggestion is to stipulate that the relevant expected future credences should be only those which your future self has when he is 'sound of mind'. Surely, when you are intoxicated all bets are off. Yet, being sound of mind cannot simply mean being rational, otherwise the principle would read: it is part of rationality that if one believes that one's rational credence about some proposition equals  $x$  at some later time, it should equal  $x$  at present. This is not a very illuminating principle: the reasons I can have for thinking that the future credence is rational are reasons that inform my current credence: if I have reasons that defeat my current credence, I am synchronically incoherent, and if I have reasons to support my current credence, the expected future credences are redundant.

The definition of soundness of mind may be a medical problem. Yet, as long as the medical definition is not synonymous to rationality, this construal leaves open a loophole for scenarios of a medically sound future self whom I nonetheless expect to adopt a credence that I could not endorse currently. I believe that the probability that Michaëlla Krajicek wins the next Wimbledon is quite low, but I also believe that when I will watch her upcoming match, I will get excited and genuinely overestimate her chances. Although *I* believe that my future judgment is, in some sense, clouded, this is no medical case of unsoundness of mind. The hypothetical psychologist might analyse me as suffering from sudden overexcitement, but I might, just as well, suffer from a chronic pessimism which is only temporarily relieved by the excitement. The point is this: my expected future credences only have impact insofar as *I* can endorse them *at present*. Thus, the 'direction of import' seems to be the other way around: it is not the case that my expected future credences should inform my current credences, but my current credences should (and do) inform my expectations.

Van Fraassen discusses these problems under the apt header 'the problem of Ulysses'—after Homer's hero who anticipates the sirens and distrusts his future beliefs. In response, Van Fraassen indeed puts a bridle on the future

expectations to which one should commit: “to follow only epistemic policies which *you* can endorse.” (1995b, p25) Hence, my current credence should equal my expected future credence iff I could currently endorse it.<sup>16</sup> This admittedly copes with cases of foreseeable irrationality, but since ‘endorsement’ here must mean *rational* endorsement, reflection can no longer function as an illuminating, normative principle of rationality. As Foley argues, any agreement between current credences and the endorsed expected future credences derives from the reasons I have to endorse my current credences. (2001, p165) Reflection, then, reduces to the advice to aim for relatively stable beliefs. The view of belief as engagement pertains to one’s integrity as an epistemic agent: to adopt a belief involves a future-directed decision. The refusal of this doxastic engagement is pragmatically incoherent in Moore’s sense. (Van Fraassen 1995b, p26; 2004, p173)

### 1.3.3 Almost Anything Goes

At first blush, the requirement of coherence might offer the voluntarist some resources to put a further bridle on irrationality. A belief-web typically includes an implicit epistemic policy that governs its credence kinematics. (Van Fraassen 1980a, p165–8) It includes, *inter alia*, what one takes as (sources of) evidence (experience, explanatory power), the kind of inferences one is willing to make and certain values which may help in balancing out James’s opposite epistemic desiderata (which beliefs are sufficiently important to me in order to risk embracing falsehoods, how much information do I desire to engage epistemically). We are committed to uphold both this policy and the further beliefs which are formed under its governance.

Given Van Fraassen’s portrayal of the human condition as Neurath’s sailors repairing their boat at open sea, we cannot just believe anything, but we are to a large extent constrained by our prior opinions, i.e. we are and always have been and always will be on a boat. (Van Fraassen 2000, p279) The label ‘voluntarism’ notwithstanding, we are shaped by our ‘mother’s knee’-beliefs and we cannot voluntarily adopt any (arbitrary) outlook. The ‘historical understanding of the self’ and the broad notion of coherence, thus, put some plausible restraints on epistemic behaviour. Yet, the inclusion of prior ‘mother’s knee beliefs’ is *descriptive*. As a matter of fact we happen to share a significant body of beliefs with one another. This observation, however, does not render the devious madman irrational... only different.

The appeal to Neurath’s boat is intended to show that *not anything goes*, and is explicated by yet another Jamesian notion: *live options*.<sup>17</sup> For

<sup>16</sup>Van Fraassen (1995b) also generalizes the principle such that my current credence should fall within the range marked by my expected future credences that I could currently endorse, thereby accommodating both indeterminate credences and conflicting expectations.

<sup>17</sup>Also see Teller (2011, p64) on homeopathy and regular medicine; although both sys-

instance, the thesis that safety does not require fastening your seat belt is not a live option; it is hard for us to imagine how we could come sincerely and coherently to believe such a thing. Thus, Van Fraassen remarks concerning person A, who entertains this thesis:

I would still judge A to be irrational even if I thought his opinion was coherent. My explanation would consist in *my selecting* certain of his opinions as ‘his evidence’, and evaluating how his opinions go beyond *those* by my own standards for prudent extrapolation and risk assessment. (1992, p25)

Yet, this is strange! Voluntarism only demands coherence, and this is an internalist constraint in the sense that whether or not someone’s belief-web satisfies it, depends solely on this belief-web. The coherence does not need to be explicitly acknowledged by the believer. His belief-web is, in principle, accessible by him and his rationality is not affected by a comparison to anything outside of this web. (1985, p248) In addition, voluntarism denies a foundational role for evidence. The best way to reconcile this, I think, is to assume that Van Fraassen is temporarily evaluating *from his personal perspective*. Then, the term ‘irrational’ is ill-chosen by his own standard. Rather, he should have classified it as e.g. ridiculous.

This interpretation is indeed suggested by a comparable example (by Ladyman 2004, p142): the thesis that shooting someone in the head alleviates migraine is not a live option; we call it absurd. But Van Fraassen hastens to add, “I do not classify you automatically as strictly speaking irrational on the basis of your medical beliefs alone – but is that much of an objection?” (Van Fraassen 2004, p184) The pejorative classification is not mandated by voluntarism; we (you and I) just happen to share this attitude. The alleged irrationality is not grounded in prescription, but in a description of our shared but optional opinions. If we disapprove of someone for trying to alleviate his migraine in this way, we can only appeal to our own beliefs and values, and these personal matters are no part of a strictly epistemic account of rationality. Thus, we have travelled a long way from a conceptual analysis of rationality to a (committed) boo/hurray-account of

---

tems have, if coherent, an equal claim to rationality, it would not be rational *for us* spontaneously to prefer a folk remedy for cancer, so Teller says. The reference to ‘our’ tradition is important here and ‘spontaneous’ should mean ‘incoherent with that tradition’ if it is indeed an irrational preference. Foley struggles with the same problem in developing his internalist, ‘egocentric’ theory of rationality. He slides from the cautious observations that “some might be rational albeit fundamentally misguided” and “it may be that [...] in general we are neither dogmatic nor thoroughly misguided” to the less cautious “contingent fact that we are born with similar cognitive equipment and into similar environments [...] that makes it likely that the deep epistemic standards of one person will not be radically different from those of another.” (Foley 1990, p80–1) Whereas Van Fraassen and Teller appeal to tradition, Foley, thus, appeals to physiology and environment. Both are non-normative: as far as rationality is concerned, our coherent beliefs cannot be reproached or ratified, only condoned.

epistemic evaluations—that is an objection indeed. That is, unless it is an irrevocable and accurate analysis of the human condition; then we had better learn to live with it. But there is something more to say about seat belts and migraine without thereby sliding into either defensive epistemology or debilitating relativism.

## 1.4 Reconceptualising Rationality

### 1.4.1 Standard of Appraisal

An account of rationality lays down a standard of epistemic appraisal. This is best illustrated by revisiting the norm of probabilistic coherence. Since the late sixties, a body of psychological, empirical knowledge started to emerge which shows that actual humans often violate the probability calculus. To a certain extent, these experimental findings can be given an alternative interpretation, such that the test subjects misunderstood the questionnaire, but were perfectly coherent in their answers *given* the way they understood the question. The empirical findings also prompted research programmes in evolutionary psychology, claiming, for instance, that our ‘heuristics and biases’ or rough and ready rules of thumb may not be strictly coherent, but are rational nonetheless, since they generally promote survival. The first strategy explains away the alleged incoherence, while the latter accepts the incoherence but contests the alleged irrationality. Concerning the first strategy, it is fair to say that, as the puzzles become more difficult, real people make real, probabilistic mistakes—after all, we are not all mathematicians. Moreover, we are confronted with such puzzles in real life.<sup>18</sup>

If the constraint of coherence renders us all irrational, this suggests to some that the demand is too strong. On the other hand, the constraint can be interpreted as an ideal, such that the credences of an *ideal* epistemic agent are an interpretation of the calculus and we should all strive to be this agent. This demand can be compared to the norm of consistency: I do not oversee all logical implications of my beliefs and if I did, I would, perhaps, discover inconsistencies. I dislike thinking of myself as irrational, but the norm of consistency provides for *the possibility of reproach*. You can blame me for inconsistency (praise me for consistency) and if you point it out to me, I revise my belief-system. The proper function of probabilistic coherence is to ground our intersubjective, epistemic evaluations.

Foley argues, against this conception, that we cannot be blamed for deviating from an unattainable ideal. “Real human beings can make reasonable mistakes even about matters of logic and probability, and any account of

---

<sup>18</sup>This type of research was pioneered by Tversky and Kahneman. For examples of the decision tasks that were designed, an overview of the experimental findings and the various frameworks for interpretation, see e.g. Samuels and Stich (2004). Their conclusion is that we perform perfectly well on certain tasks, while making systemic errors on others.

rational belief or decision that implies otherwise must be dismissed as inadequate.” (1992, p185) He allows for reasonable mistakes; rational beliefs that nevertheless violate some other standard, namely the standard of logic. Yet, if even logical defects are no solid basis for epistemic appraisal, I guess discussion comes to a halt. Alice points out that Bob has a contradictory pair of beliefs. Bob retorts that this is fine for rationality requires no non-contradiction. How should Alice respond? She might say that the doxastic aim is to believe all and only truths and that these contradictory beliefs cannot be jointly true. Bob may easily contest this; he has already rejected non-contradiction. It is hard to justify logic in non-circular fashion.

Perhaps Foley’s ready dismissal is prompted by unease with a standard that renders everyone, strictly speaking, irrational. If so, this misinterprets the function of a standard of appraisal. ‘Rational’ and ‘irrational’ are, figurative, not quantitative, extremes, although the terms are used slightly differently: I can praise one for being rational in specific circumstances without implying she has no epistemic defects whatsoever, but irrationality is a stark accusation. Do I claim that both rain and not rain are highly likely? You say I am confused. Did I make a mistake in my bankbook? Sloppy. Am I in mortal agony for testing probably positive on some extremely rare disease? You say I overreact. Praise and blame come in many flavours, but *to be on target*, they should appeal to the epistemic standard of rationality.

### 1.4.2 Evidential Effort

There is an additional norm that cuts across all epistemic policies, i.e. that of *evidential effort*. We formulate it carefully, so as to not slide into an untenable defensive position and at the same time make the case that it *extends* voluntarism. First note that we have no *direct* doxastic control: we cannot just decide to dis/believe sincerely any *p*. However, we have *indirect* control, in the sense that we have some control over our informational position. We can indirectly change our beliefs by actively controlling our informational positioning, e.g. by consulting an expert or the encyclopaedia, by performing experiments or simply going some place and observe. Although a practicing scientist might have a hard time denying the recalcitrant phenomenon in front of him, he might avoid anomalies in a more roundabout way: by not placing himself in the position whence potentially conflicting information is acquired. The upshot: we wish to propose an epistemic norm; ought implies *can*, and we *can* alter our informational circumstances, so the norm is at least not precluded by contraposition.

An account of rationality guides us in reaching the two-fold epistemic aim. We simultaneously try to gather truths and avoid falsehoods. Thus, the activity centrally involves the acquisition of *information*: propositions that are believed (to be true). To believe anything means to expose oneself to the risk of believing a falsehood. Regardless of one’s implicit epistemic

policy, this risk may be reduced by acquiring more information, in the sense that a false belief can be debunked by further investigation. I take it this connects with the voluntarist point that coherence with one's current body of beliefs is, as a matter of fact, a strong constraint. By not placing ourselves deliberately in an informational vacuum, we expand our web of beliefs, and when potentially conflicting information turns out to cohere, we can say we have harvested (corroborating) evidence. Even if the balancing of the epistemic desiderata is something that cannot be captured in a protocol, there might be a genuine *third way*, which does not assert ampliative rules of inference, but does increase the normative space of epistemic reproach.

Similar intuitions emerge in critical appraisals of voluntarism. Alseptor-Kelly most recently put it in a nutshell: “voluntarism is so wildly permissive that it countenances as rational belief-sets that are obviously completely crazy, including belief-sets which completely disregard all empirical evidence.” (2012, p189) Psillos presents an extended criticism on voluntarism's failure to accommodate evidence.

If we are *reflective* about the content of our beliefs, then we need to consider what evidence supports them and not just how they cohere with the rest of what we believe. If we are epistemically *responsible* [...] we should consider how we should conduct inquiry in a way that secures that our beliefs bear the weight of the evidence. If we want our beliefs to be *responsive to reasons*, then again we should examine how the relevant evidence supports the beliefs we have. (2007b, p147, my emphasis)

Of the three norms mentioned by Psillos—reflectiveness, responsibility and responsiveness to reasons—our norm of evidential effort seems to fall under the purview of epistemic responsibility. Now, I am not sure how our ‘beliefs can bear the evidence’, but if this was intended reversely, so that we should allocate our beliefs in proportion to the evidence, then it reminds us of the un-voluntarist endeavours to ascertain an objective warranting-relation. If we were to try to quantify it, we would end up in the desert of ampliative recipes. I return to this problem after examining the other two norms.

Reflectiveness means that our beliefs should not just cohere but be supported by evidence. How should this be spelled out? If evidence signifies whatever someone accepts as evidence, this is already covered by coherence. It might signify what somebody else deems to be evidence, but if the evaluated agent does not accept this as evidence, we are in need of a Supreme Court. That brings us back to the foundational myth of evidence as unequivocal and undeniable bedrock. The same applies to the responsiveness to reasons. Reasons are typically part of a dialogue, so let us consider one between Judge and Agent.

Agent should be responsive to reasons, but the pivotal question is: *by whose lights?* If Judge and Agent have a different epistemic outlook, they might not agree on what evidence is. Judge reproaches Agent for not being

responsive to something that Agent does not deem worthy of response, for it does not count as a reason for him. (Imagine Judge to be a madman, in which case you would equally ignore his ‘reasoning’.) Maybe Agent should only be responsive if it is an objectively good reason. How are they to decide? *Ex hypothesi*, Agent does not believe this. And if we go externalist, the norm becomes effectively useless.

If Judge and Agent happen to be part of the same tradition and share their notion of evidence, Judge’s reasons become effective. The recognition of Judge’s statements *as* reasons is grounded in intersubjective agreement. In the end, their effectiveness derives from the fact that Agent endorses them. For instance, if Judge points out that some belief that Agent cites as evidence is actually irrelevant to the thesis and Agent accepts it, then it is apparently somehow part of his policy, e.g. to take Judge as a reliable source of information, and this is covered by the norm of coherence (put bluntly: Agent cannot coherently consider Judge to be a trustworthy expert and ignore his judgments).

Yet, suppose that Agent trusts Judge’s judgments and that Judge points out an informational opportunity. Our norm of evidential effort might hold that Agent should explore the opportunity. This, I claim, is *not* covered by the norm of coherence. Coherence involves accordance with one’s own policy, but this policy expresses one’s doxastic values and epistemic decisions, including decisions on when some belief is sufficiently supported. (Some people may be satisfied easily and entertain a lazy policy, although they would not refer to it in these terms.) It is not incoherent—no self-sabotage by your own lights—to hold that you have enough information to maintain an epistemic commitment, while recognizing that you could be in a better informational position. This is plausibly the case for a lot of our beliefs.

The norm of evidential effort is, thus, internalist, as it relates to informational opportunities *by one’s own lights*, but it is not covered by coherence: it is an *addition* to voluntarism. The dialogue is a dramatic device to illustrate a state in which you come to believe to have a relevant informational opportunity, but there is no conversation needed to arrive in such a state.

How shall the norm of evidential effort be qualified? We want it to be effective whenever somebody is denying the presence of the elephant in the room, but we can hardly prescribe, like Clifford, that “it is wrong always, everywhere, and for anyone, to believe anything upon insufficient evidence.” (1879, p186) Contrary to Clifford’s norm, evidential effort does not concern warrant or support, but informational positioning.<sup>19</sup> Recall the concept of

---

<sup>19</sup> At this point, we could relativize ‘sufficiency’ to the believer’s judgments on sufficient effort. Foley proposes to make sufficiency a function of value and prospective improvement: “As the stakes of my being right go up and as the chances for improving my epistemic situation with respect to the issue go up, it is reasonable for me to increase my efforts.” (2005, p320) Yet, I doubt such a ‘function’ can provide more than the illusion of quantification. More importantly, it would be subsumed under coherence, and still allows for

rationality as a regulative ideal; a standard of appraisal, providing a normative space of intersubjective epistemic reproach. Just as we can always and everywhere be reproached on the basis of coherence, I suggest we can always be reproached on the basis of evidential effort, with the only difference that the informational opportunity should be recognized by the agent. Thus, if she deems the opportunity irrelevant or non-informational, the reproach is off target. Whenever it concerns the option of acquiring evidence *by the agent's lights*, she may be reproached (in many flavours). Of course, she may retort that this opportunity has no priority for her, that she does not have the time or the resources to explore it, and so on, but these are but pragmatic excuses, not essentially related to the epistemic aim.

In sum: the norm of evidential effort prescribes that the agent should always strive to be in the best evidential position by her own lights. Evidence is interpreted as beliefs which she deems relevant to, e.g. possibly in conflict with, some other belief. By making the notion of informational opportunity *perspectival*, so that the agent must recognize it, the general norm of effort cuts across different epistemic policies, for it relates only to the epistemic aim. The norm is a genuine addition to voluntarism, since it is not covered by coherence. It extends the normative space of epistemic reproach.

### 1.4.3 Against Evidential Invulnerability

The norm of aiming for the *evidential vulnerability* of one's ampliative beliefs (the beliefs that are underdetermined by the evidence) goes hand in hand with the norm of evidential effort. An ampliative belief or body of ampliative beliefs  $B$  is evidentially *invulnerable* by the believer's lights, if the believer can foresee no possible information that could conflict with  $B$ , while recognizing that  $B$  is not entailed by his evidence. This means that the believer cannot anticipate any opportunity, not even in the distant future, that he would count as genuinely informational and relevant to  $B$ . The adoption of such beliefs, I claim, is an epistemic vice.

This clearly builds on aforementioned pointless epistemic risk argument, although 'pointless' may be a slight misnomer; aside from pragmatic considerations, there is an epistemic 'point' to believing, namely the hope or faith that one believes a truth. Rather, we accuse the invulnerable believer of reckless or irresponsible epistemic behaviour.

*In abstracto*, we can characterize the theory-evidence-relation as a relation between beliefs and more beliefs. We shy away from the attempt at explicating a relation of support<sup>20</sup>, but the inverse relation is easy: probabilistic

---

epistemic laziness.

<sup>20</sup>The rejection of an objective warranting-relation uniquely determining evidential support and rational response to new information does not mean that nothing more can be said about the theory-evidence relation. There is something wrong with the claim that there is a mouse in the wainscoting based on the memory of your first kiss. I take it this

incoherence, including, in the case of the extreme degrees, inconsistency. For ease of exposition, we focus on the extreme case, although this admittedly results in a sketchy and incomplete characterization of our epistemic states.

Three caveats: *first*, we take both theory and evidence as (collections of) beliefs, but there is really no such clear-cut division in an agent's doxastic household. Some beliefs are presented as evidence for others, but this direction can be reversed; there may be circles, zigzags, and so on. *Second*, we reject foundational beliefs, but, although some beliefs are fairly easy to give up on, others, we might call them more basic, are tightly embraced—we can imagine 'spheres of vulnerability'. *Third*, a certain amount of dogmatism is *not reproachable*. Open-mindedness and a critical attitude are laudable, such that even the most tightly held beliefs are not indestructible, but there comes a point at which one might say, echoing Wittgenstein ([1953] 2009, 217), 'I can offer you no more reasons, this is just the way it is, this is where my spade is turned.' In my analysis, this option is accommodated by beliefs that the agent does not consider to 'go beyond the evidence.'

The reproachable feature of evidential invulnerability applies, thus, to beliefs that an agent considers *ampliative*, i.e. Underdetermined by his evidence. Let  $x$  and  $y$  be bodies of propositions or beliefs and let  $Und(x, y)$  express the relation that  $x$  is consistent with, but not entailed by  $y$ :  $x \wedge y \not\vdash \perp$  and  $y \not\vdash x$ . Let  $B$  be a belief or a collection of beliefs of some agent and let  $E$  be his evidence; a collection of beliefs of the same agent, comprised of those beliefs which he deems relevant to  $B$ .  $E$  might include the initial explanandum of an abduction to  $B$ , reasons he raises in dialogue, potential defeaters which turned out to be consistent with  $B$  (corroborators); the perceived relation of support is up to the agent.  $Und(B, E)$  then expresses that his beliefs  $B$  are consistent with—but not entailed by his evidence  $E$ . Beliefs change over time and since both  $B$  and  $E$  represent selections of the agent's beliefs, we could give them a timestamp, after all  $Und(B_t, E_t)$  might not equal  $Und(B_t, E_{t+1})$  or  $Und(B_{t+1}, E_t)$ , but we shall consider a single moment.

Our epistemic policies typically include beliefs about the *range of permissible evidence*. If we understand the problem of induction as the problem of justifying inferences from past and present experience to future experience, or from the observed to the unobserved, this can be subsumed as a claim of underdetermination, such that no unfalsified theory  $T$  about future (beliefs about) experience is entailed by any *permissible* body of beliefs about the evidence  $Q$ , i.e.  $Und(T, Q)$ . In a sense, this claim is about the range of permissible evidence:  $Q$  is intended to include past and present experience and excludes e.g. crystal balls.

It is hard to pin down what an *evidential policy* amounts to. If someone offers me the opportunity of talking to her bottled genie, I am not inclined to

---

is also what Psillos's proposal of *reflectiveness* connects with. An encompassing account of scientific rationality might be able to put a bridle on such non-support.

accept this as a genuinely informational opportunity (not in the sense that it might drastically alter my outlook). My dreams have no predictive import. I rely on experience, and quantified or qualified testimony, including expertise. Such policies are not carved in stone—they are typically not explicit at all. I can imagine farfetched scenarios in which I might revise the abovementioned, but also more earthly and nuanced changes in my policy.

Even if we keep the implicit policy constant, it allows for drastic changes in beliefs that are not directly relevant to the policy. I might learn from experience or testimony that strawberries are unhealthy or that the Mona Lisa is a forgery. Some beliefs are more firmly in place. The protagonists of a comic book I read in my youth sometimes visited ‘Far Away Island’ and other times ‘Timbuktu’. It took a while until somebody informed me that Timbuktu is a real place. My Timbuktu-endorsement is now firmly upheld. Although primarily supported by hearsay, the regress to Timbuktu-denial is no live option for me. Yet, within the confines of my policy, there may be permissible scenarios that unveil the Timbuktu-conspiracy and make me a dupe, although part of believing is, of course, to expect that no conflicting information will present itself. Moreover, the permissible evidence is not limited to information I expect to be obtainable during my lifetime. The discovery of extra-terrestrial life might happen in the remote future, but the relevant evidence—plausibly expert testimony—is ratified by my policy.

Let  $P$  denote the admittedly vague and messy mixture of down to earth and fantastic beliefs that the agent at some time slice deems permissible evidence. This includes a wild bunch, but only those beliefs that could be ratified by his policy. The complement of  $P$  is reserved for the pseudo-information springing from e.g. dreams and genies. For a consistent agent, the beliefs  $E$  that are presented as evidence for any selection of his beliefs are a subset of the beliefs that are permitted by his policy:  $E \subset P$ .

Suppose an agent, let us call him Nicod, believes that all ravens are black ( $B$ ), and that the only permissible evidence  $P$  is comprised of singular experiential beliefs of the form, ‘Lo, here and now! I observe  $z$ .’ Note two interesting features of his epistemic state. *First*, if he is consistent, then  $Und(B, E)$ . If he had observed a non-black raven, his beliefs are inconsistent and his *universal* claim  $B$  cannot be entailed by a subset  $E$  of his permitted evidence  $P$ , since this consists of *singular* beliefs only. *Second*, it is not the case that  $Und(B, P)$ , since the observation of a non-black raven is among his permissible evidence and inconsistent with  $B$ . Let us call an ampliative belief or body of ampliative beliefs  $A$  *evidentially vulnerable* whenever it is not underdetermined by an agent’s permissible evidence: not- $Und(A, P)$ .<sup>21</sup>

<sup>21</sup>The criterion of evidential vulnerability reminds us of Popper’s criterion of falsifiability. The ampliative belief  $A$  is vulnerable iff either  $P \wedge A \vdash \perp$  or  $P \vdash A$ . In the latter case,  $A$  is verifiable, but this will rarely hold, because science typically deals with general propositions. In that case,  $A$  is not entailed by the permissible evidence,  $P \not\vdash A$ , and  $A$  is vulnerable only if it is inconsistent with  $P$ , i.e.  $P \wedge A \vDash \perp$ , i.e.  $\exists \varphi \in P : A \wedge \varphi \vdash \perp$ , i.e.

This means that this belief is deemed to go beyond his current evidence, but that the agent can see how it might be either established or rejected by evidence that is allowed under his current policy.

If Nicod were to observe a non-black raven, he comes to believe that not- $B$ . This belief is entailed by his evidence; it is not undetermined. Of course, we could ask him for evidence for his evidential beliefs, and he might respond by giving further reasons why his observations can be trusted, *but* there is a point at which he could admit that this is where his spade is turned. He does not recognize not- $B$ , or equivalently: his observation of the non-black raven, as going beyond his evidence. By his lights, the belief is established, and not held tentatively as vulnerable to whatever the future may unveil. As intimated, a certain amount of such ‘dogmatism’, whether construed probabilistically or binary, is not reproachable. On the other hand, it may not be *ipso facto* laudable. Dogmaticus believes that all ravens are black ( $B$ ), based on his believing so  $E$ , and the only permissible evidence  $P$  is comprised by his current beliefs. His belief is ‘vulnerable’ in the sense that not- $Und(B, P)$ , but for Dogmaticus it does not go beyond his current evidence: not- $Und(B, E)$ . Since we are in the process of analysing the reproachable state in which an agent deems a belief to be *not* established, *nor* corrigible by permissible evidence, we grant Dogmaticus his dogma; if defective at all, it should be diagnosed as a different defect.

The vice we condemn is exemplified by Reckless, who is like Dogmaticus, with the exception that his universal raven-belief is based on his observations of one black raven and a white shoe  $E$ , thus  $Und(B, E)$ , and, since  $P$  is comprised of his current beliefs only, we see that  $Und(B, P)$ . This means that Reckless recognizes that his belief is not self-evident and goes beyond his evidence, while seeing no way in which this claim could be affected by future information.

Why is Reckless reckless? The norms of evidential effort and evidential vulnerability are tacitly motivated by the assumption that false beliefs can be debunked by gathering more information, and this, in turn, only flies if the subject of investigation offers ‘resistance’. If the doxastic target—typically: the world—cannot offer friction, our beliefs about it are *gratuitous*. An implicit epistemic policy includes an outlook on evidence. ‘Evidence’ and ‘information’ are success-terms, and part of the concepts of evidence and information is that it is not gratuitous. In Reckless’s eyes, the target of his ampliative belief cannot offer any additional resistance; he cannot be reproached on the basis of epistemic effort and has, thereby, withdrawn himself from the realm of intersubjective epistemic appraisal. Therefore, his belief is not only gratuitous, but incorrigible: its potential falsity cannot be exposed. There is a curious dogmatism involved in the recognition that one’s belief might very well be false, but that it cannot, not even in principle, be

---

$A \vdash \neg\varphi$ , in which case the criteria coincide.

debunked. This recklessness is reproachable by our epistemic standard.

As intimated, this analysis is sketchy and incomplete. A more encompassing account should render the above in terms of credences, so that the binary entailment and inconsistency are replaced by probabilistic coherence and ampliative beliefs are probabilistically invulnerable if the agent permits no evidence that would affect his credence-assignment.

The norms of evidential effort and evidential vulnerability enrich the base-line of probabilistic coherence, put a further bridle on irrationality and enhance the normative space of epistemic appraisal, while offering sufficient room for personal, value-driven decisions and rational disagreement, also in the light of equal evidence. Together, they give a partial account of *epistemic responsibility* or doxastic prudence. Evidential vulnerability effects that an agent is reproachable whenever he adopts a belief that, by his own lights, goes beyond the current evidence, while his evidential policy permits no additional evidence that could expose the belief as false. Evidential effort requires putting himself in the best informational position attainable—again, by his own lights. Epistemic responsibility encourages open-mindedness and critical investigation and combats dogmatism and wishful thinking. It depends solely on the abstract, two-fold epistemic aim and, consequently, cuts across rival epistemic policies, i.e. it is ‘common-stansical’.

Responsible information-oriented activity does not, by itself, get us out of relativism. Still, the notion of information presupposes that the subject of investigation is able to offer resistance, in which case not anything goes. If this resistance is such that it leads to long-run, intersubjective agreement on the evidence, we might move closer to each other and render epistemic ‘boo’s’ and ‘hoorah’s’ less arbitrary.

## 1.5 The Scientific Context

### 1.5.1 Empirical Evidence

The opening dialectic deadlock demonstrates that the most important arguments in support of CE (making sense of science, no metaphysics, no pointless epistemic risks) rely on a prior epistemic outlook which, in turn, stems from the empiricist stance. Taking these two ingredients as a package deal, the arguments are circular. Voluntarism renders any coherent package rationally permissible. On the bright side, we did, *en passant*, manage to find some potentially shared, common-stansical commitments, namely the recognition that experience has some ‘imposing force’ and that the realist may agree that scientific theories should, at least in the long run, be constrained by experiential evidence. These ideas may be put to work to relax the deadlock.

The emphasis on (sensory) experience comes down to an endorsement of the intuitively appealing view that experience is the long-run final arbiter on

scientific theorizing. This endorsement is widely shared: most serious views of science, save the more radical ‘strong’ programmes in social constructivism holding that scientific theorizing is mostly constrained by social factors, involve such an empirically oriented engagement, if only as a methodological cue to actively and critically subject our theories to experiment. Given the potential agreement on evidential considerations and the emphasis on experience, it seems a missed opportunity that Van Fraassen does not sow this common ground. How, then, does he envisage the role of empirical evidence?

While trying to sidestep subtly Feyerabend’s criticism on foundationalist empiricism, Van Fraassen (2002, p117–43) allows the empiricist to get some grip on experience. Feyerabend compares the purported *primacy of* experience to the alleged authority of the Holy Scripture. If we take the Scripture to be our sole source of information, three questions emerge: What is it? What does it mean? What does it imply? We cannot distinguish authentic from apocryphal on the basis of these texts themselves, and we cannot, *ex hypothesi*, appeal to some extra-Scriptural source of information. Even if we could, we might have alternative interpretations and these are not settled by citation. Finally, identification and interpretation do not determine which further consequences to draw if the information is, in some sense, incomplete.

The parallel with experience is obvious. We cannot identify it on the basis of itself, we have conflicting theoretical interpretations, and it leaves much underdetermined without providing an inbuilt ampliative algorithm. Without such foundations, it seems, we are faced with either dogmatism or scepticism.

However, Van Fraassen says, reliance on the Scripture *in actual practice* has proved less problematic than this anti-foundationalist criticism makes it appear. Pious communities fare well by drawing on *tacit understanding*: a tradition providing for more or less tentative answers to the three questions. This tacit understanding will not yield many cross-communal converts; once it is made explicit, the problems simply iterate. But *from within* the tradition it provides some self-affirming guidance.

Hence, non-foundationalist reliance on experience is purportedly enabled by an *undogmatic engagement to communal understanding*. Is this understanding to be equated with the implicit epistemic policies that come along with taking an epistemological stance? Then, the ‘empiricist community’ would provide implicit answers to Feyerabend’s questions. Van Fraassen mainly mentions examples of *personal understanding*. Thus, whether some alleged experience was veridical or not (an oasis or a mirage), whether I saw a zebra or a painted mule and what further conclusions I draw, is in principle always up for debate, but while acknowledging my fallibility, I ordinarily and implicitly consider myself to form new—and proceed from—reliable opinions. Yet, these are no encompassing answers, but commonsensical examples of personal case-by-case judgments.

Let us, by contrast, consider some encompassing answers to Feyerabend’s

three questions. One might, for instance, claim that revelation or intuition should be identified as authentically experiential—and are, thus, *a posteriori* vulnerable to conflicting revelations. The scientist behind the microscope may hold that he looks *through* the instrument and directly observes the cell (cf. Alspector-Kelly 2004). Even if we interpret observation as unaided, beliefs about the cell might be arrived at by the friend of explanatory loveliness (cf. Hacking 1985).

What would be a rival understanding of the empiricist community? Perhaps the direct realism that asserts, roughly, that we have veridical experience of medium-sized objects. How to *identify* authentic (i.e. veridical) experience? A causal story, e.g. about the physical processes involved in visual registration, would help here, but this is, of course, no story the empiricist is happy to tell. What about a procedural characterisation along these lines: close your eyes and if it is gone, it was authentic (visual) experience? Now, this does not solve the *foundational* problem. If all information is supposed to come from experience and the identity of experience is at issue, we cannot appeal to more purported experience or a lack thereof, because the problem iterates. When we are in the process of identifying experience, any appeal to experience is circular. Yet, such guidelines may be helpful in those rare cases that the identification problem arises *in practice*.

Regarding *interpretation*, the empiricist may want to minimize the overlay of theory. Again, a causal story would be convenient. For instance, Buekens and Muller (2012) distinguish, *inter alia*, between registration, ‘Object-seeing’ and ‘Doxastic seeing’, where only the latter involves interpretation: observing *that*  $X$  is an  $F$ . These distinctions might help to combat theory-infection. Alas, the empiricist cannot make the difference between seeing and seeing *that* solely based on experience (although he can make the difference between observing *that*  $X$  is an  $F$  and observing *that*  $X$  is a  $G$ ). A human organism can be considered a measuring device, but a measurement “shows not how the phenomena *are* but how they *look*.” (Van Fraassen 2008, p290) Van Fraassen’s direct realism might save him from *phenomenalism* or *idealism*, but he cannot rely on the supra-empirical aspects of scientific accounts of vision and detection to tackle the interpretation problem.<sup>22</sup>

Schurz’s approach to minimizing theoretical overlay (Schurz, unpublished) is better suited to the empiricist spirit. He proposes to ‘measure’ grades of interpretation by intersubjective, ostensive learning curves: the steeper the curve, the more theory-infected the concept is. The flatter the curve, the less theory involved, and the closer the concept comes to being a ‘neutral observation concept’. Like the pragmatic ‘close your eyes’-reply to the iden-

<sup>22</sup>This becomes more pressing when we notice that the scientific account of vision tells a similar story about technological enhancements (in which case we may have good reasons to accept other instruments as members of our epistemic community and rely on their expert testimony). Indeed, epistemology is in dire need of a stable analysis of the ‘Janus-faced’ character of experience—with its ‘interpretative’ and its ‘imposing’ face.

tification problem, Schurz's method does not solve the *foundational* problem, because it relies on *empirical* experiments, which are in need of interpretation themselves. Yet, proceeding from a shared understanding of experience, his method does address interpretation *in practice*.

Finally, assuming that the identification and interpretation of experience are settled, the empiricist holds that conclusions from experience that go beyond what we can experience are illegitimate. (see Chapter 3)

Still, these proposed understandings, flowing from empiricist tradition, offer no secure foundations. If these guidelines for the proper identification and interpretation of experience are based on experience themselves, then they are circular. If they are based on some alleged extra-experiential source of information, then they are inadmissible. Moreover, if another source of information is assumed, Feyerabend's questions readily return to haunt this new foundation. The criticism on epistemological foundationalism is very general; it applies to any claims about a source of information.

Notice however, that there is a good reason that Van Fraassen had to revert to case-by-case examples to illustrate our implicit understanding in practice. There is an important disanalogy between experience and Scripture that he failed to make explicit. With Scripture there is a relatively stable body of texts which need to be authenticated, but once this is done and agreed upon within a community, the identification problem has come to a halt—except, perhaps, for those rare incidents at which some new scroll is unearthed. With experience, identification is an on-going process and judgments on authenticity are made case by case. The same goes for interpretation; was the stick bent or distorted by the water or both? Authentic experience and proper interpretation cannot be singled out beforehand.

As intimated, there is some common ground between rival views of science, relating to a scientific emphasis on experience. Although the realist might claim that there is no interesting epistemic difference between observational processes that do and do not involve the aid of instruments, the conceptual difference between (unaided) experience and inference from (unaided) experience is not elusive to him. He can also acknowledge different levels of interpretation involved in aided detection; some of his causal stories are more speculative than others, i.e. some types of 'experience' have more 'imposing force'. Finally, science, as realistically construed, has had to get off the ground somewhere. Our unaided experience provides a suitable starting point. Let me stress that I do not claim that the *hegemony* of (unaided) experience is agreed upon, but only that philosophical views of science should have something to say about (unaided) experience.

Notice further that there is an order or hierarchy involved in Feyerabend's questions. Identification and interpretation may be intimately intertwined, but implication (i.e. what conclusions we may draw) seems to be a quite different and subsequent affair. Let us temporarily bracket this third question, which is the subject of epistemological controversy.

Voluntarism relativizes the identification and interpretation of evidence to personal value-driven judgments. One's epistemic outlook is not dictated by the evidence; it is the other way around. I suggest, on the contrary, that any epistemic outlook that takes a stance on science—plausibly: any epistemic outlook full stop—needs an implicit understanding of experience. The identification of—and negotiation on the proper interpretation of experience is the bread and butter of science. The communal understanding that Van Fraassen refers to does not have to spring from the empiricist tradition, but is a matter for the whole science-endorsing community. His heroic effort to get a grip on experience is laudable, but it is a *shared problem* that cuts across different stances, and might, thus, be relegated to shared resources. Henceforward, I proceed under the fiction that (unaided) experience is the equivocal but relatively unproblematic, solid ground for long-run agreement, at least on an answer to the first of Feyerabend's questions (identification).

### 1.5.2 The Primacy of Experience

Science is our prime example of systematic epistemic investigation. In the light of our reconceptualization of rationality, we might expect it is the epitome of responsible information-oriented activity; that scientific theories are coherent and evidentially vulnerable (I assume that all scientific theories are ampliative) and the scientist strives to put himself in the best informational positioning attainable.

In the scientific context, experience is one widely agreed upon source of information. Suppressing the fundamental epistemological problems with explicating the 'Janus-faced' character of experience, we note that its *linguistic guise* (i.e. our descriptions of the phenomena) is revisable and negotiable, but more importantly, used to point towards its *wordly face*. This wordly face has an imposing force—to put it mundane: if I bump my head, I might deny or redescribe the event, but I still bump my head. Experience plays a special role in that it trumps alleged other sources, i.e. no matter how elegant or explanatory lovely a theory is, as intersubjectively recognized anomalies heap up, there comes a point at which the theory will 'crumble under persistent resistance of reality,' to paraphrase Duhem. With experience appointed as the final, long-run arbiter, there is something like a Supreme Court reinstated, which might be able to help adjudicate disputes, albeit the 'Supreme' verdicts are equivocal and debatable; they do not enforce either quick acceptance or quick kill.

If we spell out the two abstract evidence-oriented norms of rationality in terms of empirical evidence, we end up with a fair and sensible sketch of how we could conceive of *scientific rationality*. Yet, it seems this is where we must leave common-stansical ground. Could we demand of a scientific theory that it is *empirically* vulnerable, such that any theory  $T$  is not- $Und(T, XP)$ , where  $XP$  stands for empirical evidence? I am not sure. Experience is

agreed upon as *a* source of scientific evidence. A stance on science that accepts extra-empirical evidence might nonetheless agree that theories, as a whole, need to be empirically vulnerable. On the other hand, it maintains that *empirical* invulnerability does not equal *evidential* invulnerability, so it might permit theories that are unable to conflict with (or be established by) experience. This might further depend on how we conceive of theories and how we individuate them.

In general, we see at least two viable ways of defending the *hegemony of* experience. *Either*, empiricism needs a metaphysical principle after all. It can be a relatively *weak* policy on which kinds of hypotheses are admissible, in which case non-empiricist metaphysics is ruled out as inadmissible. Such a principle is not self-refuting if it simply stipulates that it is not self-applicable. (Cf. Jauernig 2007) The obvious drawback is that such a principle can be defended only by an appeal to intuitive premises or consequences (or by an appeal to values, commitments, and so on). Therefore, it is unable to convert those with rival intuitions. *Or*, instead of aiming for an encompassing epistemology, we concentrate on the scientific context and proceed from the shared agreement on experience as *one* source of evidence, to a criticism on alleged alternative sources. This is the route I take, and in the next Chapter, I show that at least one important candidate, namely explanatory loveliness, can be crossed out.

### 1.5.3 Responsible Risks

We turn to a *conditional* argument. In a nutshell: *If* there are no sources of scientific evidence other than experience, *then* realism is reckless. A scientific theory can be viewed as a specification of a family of models. Of these models we can single out the *empirical substructures*: the structures that represent the phenomena. A theory is empirically adequate iff it has a model for every phenomenon in its domain. Let  $\text{EmpAd}(T)$  be the belief that theory  $T$  is empirically adequate, let  $\text{True}(T)$  be the belief that the theory is true— various flavours of realism have various ways of qualifying, restricting and nuancing the truth-claim, but for our present purpose, this can be ignored—and let  $\text{ThAd}(T)$  be the epistemic commitment to the non-empirical substructures of  $T$ , such that  $\text{True}(T) \leftrightarrow (\text{EmpAd}(T) \wedge \text{ThAd}(T))$ . Let  $E$  be the general collection of scientifically permissible evidence and  $XP$  the empirical evidence.<sup>23</sup>

As the terms suggest, all and only the empirical substructures are vulnerable to empirical evidence, because the structures represent the phenomena and the phenomena are what we experience, such that  $\text{not-Und}(\text{EmpAd}(T),$

---

<sup>23</sup>The collection of permissible beliefs cannot be analysed solely by its form; the beliefs have to be ratified by the implicit evidential policy, i.e. the agent/s would accept it as genuine information. In the case of CE, this could be spelled out in terms of data models of the phenomena, cf. (Van Fraassen 2008, p166–72).

$XP$ ) and hence not- $Und(\text{True}(T), XP)$ , but  $Und(\text{ThAd}(T), XP)$ . If  $E = XP$ , then  $Und(\text{ThAd}(T), E)$ , which means that the belief in supra-empirical adequacy is *invulnerable* to the scientifically permissible evidence. Since  $\text{True}(T) \rightarrow \text{EmpAd}(T)$  and not- $Und(\text{EmpAd}(T), E) \rightarrow$  not- $Und(\text{True}(T), E)$ , the True-belief is not underdetermined by the permissible evidence, but we see that it is only vulnerable by encompassing the EmpAd-belief, and therefore, minimally, an ‘epistemically irrelevant conjunction’. Van Fraassen famously remarked that it is not an epistemological principle that one might as well hang for a sheep as for a lamb. (1980, p72, 1985, 254–5) We can put it somewhat more rigorous: the ThAd-belief is not only supererogatory; it is epistemically irresponsible, since its possible falsity is, *ex hypothesi*, immune to exposure. On the basis of the general account of rationality, it is therefore reproachable. If one starts out with the True-belief and then drops the irresponsible ThAd-belief, one ends up with the responsible EmpAd-belief, and this is exactly how CE interprets theory acceptance. Any stronger policy advises reckless epistemic behaviour.<sup>24</sup>

Except for clarifying the somewhat cursory remarks on lambs, sheep and pointless risks, the conditional argument pins down the pivotal point of disagreement. The realist is, of course, not irresponsible, since he will not believe that  $E = XP$ . Hence, CE needs to establish that the only permissible scientific evidence is empirical evidence, whereas the realist needs an epistemological vindication of extra-empirical information. The introduction of scientific rationality helps, I hope, to dig a cross-connection between the trenches, giving a bit more flesh to an intuitively appealing understanding of scientific ‘methodology’. If the hegemony of empirical evidence is plugged in, then stalemate is turned into checkmate. Alas, common ground only gets us so far.

---

<sup>24</sup>In his reaction to Van Dyck (2007), Van Fraassen (2007, p347) proclaims: “I am quite proud never to have relied on the [...] Argument from Underdetermination” and adds that this would not be in harmony with voluntarism. Van Dyck makes the case that Van Fraassen has never put forward the argument from underdetermination, and that this is a good thing, since it must rely on the unsound claim that empirically equivalent theories are equally supported by all possible evidence (p15), which roughly equals my claim that  $Und(B, XP) = Und(B, E)$ . Yet, Van Dyck proceeds to ascribe to CE the central claim that “the epistemic virtues will have to be empirical virtues—and empirical virtues only” based on the supposition (or ‘the value’) that supra-empirical beliefs “are not additionally vulnerable”. (p27) It is unclear to me why, then, it is a good thing that Van Fraassen has never explicitly voiced the argument: it seems a compelling argument for CE, *albeit* only effective for those with prior empiricist leanings. I agree that it is unable to break the deadlock, and, indeed,  $E = XP$  is not compelled by voluntarism, but the engaged empiricist can very effectively appeal to this self-affirming argument. Therefore, it is important for Van Fraassen to give a characterization of empirical evidence, along the lines of his characterization of empirical substructures. (cf. Van Fraassen 1980, ch3)

## 1.6 Against Manifestism

### 1.6.1 Responsible Risks Revisited

*Manifestism* is the position that science aims at manifest adequacy and acceptance of a theory involves as belief only that the theory is manifestly adequate, roughly: true about the actual *observed* (past, present, future) entities (objects, events, processes). Assuming that dinosaurs are not actually observed by our epistemic community, our scientific *belief* should concern only the skulls and bones we dig up, while the commitment to the living beasts is merely pragmatic and instrumental to manifest adequacy. Whereas an empiricist believes that there have been dinosaurs roaming the earth, the manifestist remains neutral on the issue qua belief. This toy position is concocted by Railton (1989, p234–5) as an imaginary enemy to CE and, while it emerges under various mis/nomers in many critical appraisals of CE, Van Fraassen has only commented briefly.<sup>25</sup> This response is underdeveloped and unsatisfactory; manifestism is a *serious threat*, not the least because CE's central arguments against realism count in favour of manifestism and *vice versa*.

Why is manifestism a serious threat? It holds, quite plausibly, that our epistemic access to the world is restricted to what we actually observe, in which case the permissible scientific evidence underdetermines the non-manifest complement of empirical adequacy, rendering the complement and CE *reckless*.

If I believe the theory to be [empirically adequate] and not just [manifestly adequate], my risk of being shown wrong is exactly the risk that the weaker, entailed belief will connect with actual experience. Meanwhile, by avowing the stronger belief, I place myself in the position of being able to answer more questions, of having a richer, fuller picture of the world, a wealth of opinion so to say, that I can dole out to those who wonder. But, since the extra opinion is not additionally vulnerable, the risk is—in human terms—illusory, and *therefore so is the wealth*. [...] [A]s far as the enterprise of science is concerned, belief in the [empirical adequacy] of its theories is supererogatory. (Van Fraassen 1985, p255)<sup>26</sup>

---

<sup>25</sup>Churchland (1985, p38–41) is, as far as I was able to determine, the first to challenge CE to explain the epistemological relevance of the distinction between the observed and the observable. Railton (1989) baptized 'manifestionalism', which I have abbreviated. It is further discussed by Rosen (1994, p161–2), Psillos (1999, p189–90), Ladyman (2000, p852–3; 2004, p757–8; 2007, p47–51), Alspector-Kelly (2001 p416–8; 2006, p372–4), Monton and Van Fraassen (2003), Berg-Hildebrand and Suhm (2006) and Van Fraassen (2006). Ladyman (2007, p49) has changed his label into *actualist* empiricism (with reference to Kukla) which is a misnomer, since CE and manifestism agree on the 'actual' ingredient, but not on the observed/observable ingredient.

<sup>26</sup>I have replaced 'truth' by 'empirical adequacy' and 'empirical adequacy' by 'manifest adequacy.'

If CE relies on the accusation of recklessness to reject realism, this might convince the empiricist. In the literature, we discern two lines of response: (i) manifestism omits opinions that ‘we care about’, and (ii) manifestism fails to make sense of actual scientific practice. (The second is dealt with in the next Section.)

The (i) ‘what we care about’-response is not explicitly espoused by Van Fraassen, but could be developed along voluntarist lines: “If you don’t believe something you risk the loss of a valuable, worthwhile way of seeing your own situation and the world you are in.” (2001, p168) Against the manifestist, we could claim that we value our beliefs about the non-manifest observable (including dinosaurs). Yet, as a *general* attack on manifestism, this drives us right into the arms of the realist. If we value these additional beliefs *while acknowledging* that they are empirically invulnerable, such that underdetermination by the permissible scientific evidence is allowed for, this can just as well be said of claims about the unobservable.<sup>27</sup> Note that this argument fails, *not* because realist stances have other epistemic policies, but because CE cannot condone invulnerable beliefs, however ‘valuable’, to contest manifestism.

### 1.6.2 Making Sense of Science

Appeals to a superior portrayal of scientific activity are, as mentioned, troublesome and tend to be question-begging.<sup>28</sup> Also, they are difficult to adjudicate. Manifestism does not deny that supra-manifest considerations play an important pragmatic role and CE agrees, at least *conceptually*, that experience equals actual observations.

In his review of CE, Rosen (1994, p161–2) explores manifestism and immediately adds a refutation on the motivational level: (M-M) if the manifestist is in the unique position to unearth a certain artefact, he has no scientific motive to grab his shovel, since the artefact is unobserved (past, present and future, by hypothesis) and not covered by manifest adequacy. Rosen also (p162, fn13) suggests an analogue attack on CE (CE-M) concerning the unique opportunity to experimentally transform a non-actual observable event into a phenomenon.

Monton and Van Fraassen take up both challenges. Regards (M-M) they refer to some ‘agreement’ to the effect that manifestism “is incompatible with what it is virtually universally agreed about scientific practice” and

---

<sup>27</sup>A similar ‘care-about’-argument is proposed by Nagel (2007, p38) and anticipated by Railton, who answers poetically: “If you think a lovely, wide view is worth crossing a few bridges, why isn’t a lovelier, wider view even worth crossing a few more?” (1989, p248)

<sup>28</sup>This is the reason why I do not think it is fruitful to follow the trend initiated by Rosen (1994, p162–3) to portray CE as striking the best balance between two competing desiderata: epistemic modesty and making sense of science. Although this construal is wholeheartedly endorsed in e.g. Monton and Van Fraassen (2003, p407), see Van Fraassen (2006) for nuances.

add that examples like Rosen’s buried artefact show that manifestism “fails to capture our idea of what it is to do good science.” (2003, p407) Yet, on the next page they offer an interesting, three-pronged reply to (CE-M) that can straightforwardly be usurped by manifestism:

[Firstly] in the thought experiment in question, the scientists have the choice to do the experiment, and if they choose to do so then they will be making a phenomenon actual. So [...] extra empirical strength will arise from making a correct prediction about a new actual phenomenon; there is no reason the constructive empiricist cannot be interested in empirical strength of this sort. The second [problem] is that it ignores the social dynamic of science, where competition is one of the keys to success. [...] Thirdly, it is unrealistic to think that this creation of new phenomena throws no light on the phenomena naturally occurring outside the laboratory. (*Ibid.*, p408)

In manifest terms: the scientist grabs his shovel since adequacy to the artefact improves manifest strength, potentially rendering rival theories inadequate and throwing new light on other observations. I do not see how this argument adjudicates between CE and manifestism, and the curious reference to ‘agreement’, when properly spelled out, may just be grist to the realist mill. (On a side note, I do not see why such an extensive response is needed; both CE and manifestism can just reject the thought experiments since we are plausibly never in this *unique* position *and* aware of it.)

### 1.6.3 Judging Manifestability

The best reaction, I propose, is to bite the bullet, chew a bit and spit it out. In a sense, science is surely and precisely aimed at manifest adequacy: it retrodicts past experience and predicts future observations; the scientist makes judgments on what the future will unveil. In specific cases, this results in specific expectations. Moreover, the empirically oriented scientist can make principled judgments of what is observable. (See Chapter 3 for a further elaboration.) Yet, the *general* guideline that we need to believe only in manifest adequacy instead of empirical adequacy is, in effect, *empty*.

Indeed, if we were in an epistemic position to make accurate and well-informed judgments on the delineation between the observed and the unobserved phenomena, we should restrict our epistemic commitment to the former. That would be a very desirable informational positioning indeed! It is responsible, and in concordance with the empiricist spirit, to reduce our epistemic risk conform our judgments on our epistemic access. Alas, such judgments are simply too unstable to provide for an informative aim or doxastic policy. Reliable opinions on the general un/manifest distinction will collapse into general opinions on the un/observable, while, in specific circumstances, we trivially try to predict what will actually unfold; regrettably,

this is often gazing into a misty future.<sup>29</sup>

As with un/observability, the manifestist can claim that there are clear cases and counter cases: e.g. the living dinosaur falls clearly within the non-manifest, but even with regards to Jurassic and earlier phenomena, we have no principled, general way of keeping apart the manifest from the non-manifest observable; unexpected observations occur during the lifespan of our epistemic community; surprising remnants of long-gone phenomena resurface and these may bear informative relations to the (experimentally controlled) phenomena that we have contemporary access to. *That is why* we actively search for buried artefacts and design new experiments. (To be fair, aside from a precautionous wish to predict future experience, this activity is of course driven by genuine curiosity. However, the resulting empirical coverage renders it a *healthy* curiosity.)

This does not mean that the extra-manifest commitment to empirical adequacy should be taken as, strictly speaking, pragmatic. Neither does it mean that our belief is based on a balancing of the opposite epistemic desiderata such that the empiricist tradition just happens to value opinions about the observable over opinions about the manifest and is therefore prepared to take the courageous risk of being wrong about the non-manifest. Rather, it means that the manifest dictum is *practically empty*, for it presupposes a view ‘from nowhere’. If we decide to accept a theory, we decide to believe that the theory is adequate to the phenomena in its domain—wherever and whenever.

The same applies to all those other toy positions with weaker-than-CE teleological and doxastic components, such as restricting belief to phenomena in our spatiotemporal proximity (Churchland 1985, p40; Ladyman 2007, p49), or empirical adequacy restricted to the lifespan of our epistemic community (Alspector-Kelly 2001, p416), or excluding deadly phenomena, or phenomena in the past (Alspector-Kelly 2006, p373). There is no sensible, stable position between inductive scepticism and CE. Therefore *the empiricist should endorse CE*. Does this mean that we have arrived at a compelling inductive recipe? Not at all. CE is a view of science and we are by no means obliged to accept everything that science tells us about the actual observable (just like the realist typically will not claim that we should believe all contemporary scientific theories ahead to be true). Specific theories are accepted or rejected, and belief can be tentative, partial or otherwise qualified. We are free to decide to which extent we go beyond our evidence. The point is that a general belief in the manifest adequacy of a theory is indistinguishable from the belief in empirical adequacy, since our stable judgments concern the latter only.

---

<sup>29</sup>This may be the significance of Van Fraassen’s cursory one-liner: “If we do not have a firm belief about what our future holds, our predictions must be tailored to what we think can happen, not to what we think will happen.” (Van Fraassen 2006, p144)

## 1.7 Conclusion

The terms of the realism debate should be *epistemological*. Although Van Fraassen casts it as a disagreement on how to view science, this way of setting the stage undermines the prospects of a resolution from the outset: there is a broad agreement on what activity should be classified as ‘scientific’, but both parties can successfully make sense of this activity on their own terms.

Two important arguments in defence of CE are ineffective. Insofar as the argument from ‘making better sense of science’ does *not* rely on the premise that we should get rid of inflationary metaphysics, it is not clear how CE makes *better* sense of science than realism does, although it makes equally good sense. If this argument *does* rely on an anti-metaphysical supposition, it will not convince those with anti-empiricist inclinations. The ‘pointless risk’ argument equally depends on a prior empiricist commitment. We have attempted to make the argument more precise, and although the opponent may agree on the unsoundness and recklessness of adopting or maintaining evidentially invulnerable, ampliative beliefs, there is no agreement on what counts as permissible (scientific) evidence and, consequently, no agreement on which beliefs are evidentially invulnerable. To substantiate CE, the empiricist needs to establish the hegemony of experiential evidence. To substantiate realism, the realist needs to establish an extra-experiential source of information. The debate boils down to a disagreement on our epistemic access.

The views of science and the corresponding epistemic outlooks are package deals. It is not inconsistent to combine the outlook that e.g. explanatory loveliness has evidential import with CE’s view of science, or the empiricist’s evidential policy with a view of science as an irresponsible endeavour to tell the truth across the board, but if the views rely for their plausibility on epistemological considerations, these unbalanced packages are indefensible: they have no arguments to settle the superiority of their view of science.

Voluntarism renders all coherent packages rational. Coherence may not be a weak constraint, but it does put wildly varying packages on an equal rational footing and therefore fails to break the dialectical deadlock. The voluntarist observation that people typically do not wildly vary in their opinions and epistemic outlooks, is correct, but voluntarism is unable to epistemically *endorse* this state of affairs; the overlapping consensus is, if coherent, merely *permitted*. Yet, voluntarism is presented as flowing from an analysis of man’s epistemic ‘predicament’. We have no objective warranting relation and no secure evidential bedrock. Therefore, we are free to decide on our own epistemic policy and on the risks we take in believing. Epistemic reproach is, thus, either based on coherence or on personal values.

Even if we agree with this predicament, there is more to say about rationality if we reconceptualise the *role* of rationality and portray it as a regulative ideal, opening up a normative space of intersubjective appraisal.

An account of rationality should provide a standard of epistemic reproach, where this reproach comes in various keys. I have proposed two information-oriented norms, as a stub, to work towards an account of responsible belief. If an agent entertains a belief which is—by his own lights—evidentially invulnerable (underdetermined by both his current and his permissible evidence, meaning that he accepts no opportunity where this belief can face resistance), he may be accused of recklessness. If an agent does recognize a genuine informational opportunity, but does not act on it, he may be accused of epistemic laziness. This account can be improved upon by reformulating evidential invulnerability in terms of probabilistic invulnerability, by making more precise how to think of the evidence that is permitted by some policy, and by adding, if possible, a criterion that makes personal judgments of evidential support the subject of epistemic appraisal.

This sketch of a responsible notion of rationality extends on the voluntarist norm of coherence. As it is evidence-oriented, it is suited to be applied to the scientific context. CE might profit from the condemnation of adopting invulnerable beliefs: if experience can be established as the only permissible source of scientific evidence, realism is rendered reckless. But the hegemony of experience cannot be established so easily. The weaponry is, so to speak, put in position, but the firing has to wait. Fortunately, there is a considerable agreement on experience as *a*, if not the prime, source of scientific evidence. The most promising strategy for CE is to debunk alleged alternative sources. Such a strategy cannot, once and for all, establish the hegemony of experience, but it might very well be the best an empiricist can do in good conscience.

## Chapter 2

# Miraculous Abductions

### 2.1 Introduction

*Abduction* is a way of reasoning by means of which we infer the truth, truth-likeness or likeliness of a hypothesis based on its *explanatory* connection with the evidence. For example, the fact that Harvey's theory of the circulation of the blood is able to explain why more blood passes through the heart than what the body can produce, makes the theory more likely to be true, or is a symptom of its truth. The term abduction comes from Peirce (Cf. Niiniluoto 1999). The abductive inference is nowadays often called *inference to the best explanation* (Harman 1965). In the following, I use the terms interchangeably. Inference to the best explanation has been dubbed 'the Nemesis' of scientific anti-realism, of which constructive empiricism (CE) is a species. (Kvanvig 1994, p326) We say that a theory that explains the evidence well is *explanatorily virtuous* or *lovely*. If loveliness makes a hypothesis more likely to be true, it is an *epistemic virtue*.

The no-miracles 'argument' is commonly referred to as 'the ultimate argument' for realism. (Van Fraassen 1980, p39; Musgrave 1988; Psillos 2009, ch3) In a nutshell: if we do not convert to scientific realism then the empirical and technological success of science is incomprehensible. The only or best explanation of science's empirical and technological success is realism. The no-miracles *intuition*, as I shall call it with Worrall (2011), is an *abductive* inference. To make the case that this inference is compelling, the realist needs to establish that loveliness is an epistemic virtue. The mission of this Chapter is to argue that we have no reasons to accept that loveliness is an epistemic virtue.

The literature on abduction is already a jungle and the topic deserves whole books, so I try to make my criticism on it as general as possible, thereby avoiding having to plow through the plethora of accounts of scientific explanation. Fortunately, the most elaborate defence of abduction, Lipton's *Inference to the Best Explanation* (2004), shares this premise: it is intended

as a general defence, which the explanationist (the friend of abduction) can put to work by plugging in his own favourite account of explanation. In the following, I focus on Lipton's monograph and Psillos's elaborate and long-term promotion of abduction in the context of, what could be called, his epistemically optimistic realism.

## 2.2 No Miracles

### 2.2.1 Putnam's Intuition

Abduction is a non-deductive and, consequently, semantically unsound inference: the truth of the premises does not guarantee the truth of the conclusion. If we suppose that entailment is necessary for explanation, abduction is effectively an inference to the antecedent. For instance, early in the morning, I go down-stairs and find in the hallway a wet rubber boat with some empty bottles of beer in it. I conclude that my house-mates have gone for a late-night leisure trip on the nearby creek. My reasoning may go as follows: *if* my house-mates went for a dusky boat trip  $A$ , *then* I would encounter their mess in the hallway  $C$  (this is usually the way these trips go). Thus  $A \rightarrow C$ ,  $C$ , and I *infer the antecedent*  $A$  from  $C$ , because  $A$ , if true, *best explains*  $C$ . (Henceforward, the if true-condition is omitted.)

A feature of every non-deductive, i.e. non-truth-preserving inference, is *non-monotonicity*: the inference can be defeated by adding premises. For instance, my house-mates planned a boat trip, but it was raining cats and dogs, so they decided to stay on land and use the boat as a surrogate party tent. My observation is the same: I see the wet boat in the hallway, but now I remember that it was pouring yesterday evening, and I am not so sure anymore that my house-mates would have risked wet feet: the inference is defeated by adding a premise.

In the context of the realism debate, the observation to be explained, *the explanandum*, is typically the empirical and technological success of science—various ways of making this and the other components of the inference precise are considered below. *If* science is true, *then* science is empirically and technologically successful, and since science is, indeed, empirically and technologically successful, we infer the antecedent: science is true. Truth is the *best* or only explanation of science's empirical and technological success; the staggering success is otherwise inexplicable, viz. miraculous. This is the most naive explication of the no-miracles intuition. Of course, no realist believes that 'science', as a whole (i.e. all contemporary scientific theories *aheap*), is true without qualification. Arguably the first statement of the no-miracles intuition, by Putnam, is already more nuanced.

The positive argument for realism is that it is the only philosophy that doesn't make *the success of science* a miracle. That terms in *mature*

*scientific theories typically refer* (this formulation is due to Richard Boyd), that the theories accepted in a mature science are typically *approximately true*, that the same term can refer to the same thing even when it occurs in different theories—these statements are viewed by the scientific realist not as necessary truths but as part of *the only scientific explanation* of the success of science, and hence as part of any adequate scientific description of science and its relations to its objects. (Putnam 1975, p73, my emphasis)

The several components are highlighted: science is *successful*, this can be explained by supposing that our *mature* theories typically *refer* or are *approximately true* or both, and since this is *the only explanation* of the success, its descriptive adequacy should be accepted. As it stands, the intuition is in dire need of refinement. What is it for a theory to be approximately true (Section 2.3.1)? What, exactly, is the (empirical and technological) success that is in need of explanation (Section 2.3.2)? And what may we conclude from the observation that some theory is the best or only explanation of some observation (Section 2.3.3 onwards)? These questions are explored below. For now, we leave the intuition as it is and juxtapose it with another intuition, based on the historical observation of the rejection of theories that were successful in their times.

### 2.2.2 Laudan’s List

The observations leading to the pessimistic induction over the history of scientific theory change are famously presented by Laudan (1981) (the inductive step itself is not proposed by Laudan). While reflecting on theories that enjoyed success in the past, but are at present believed to be false, he illuminates the relation between *truth*, *reference* and *success*.

Success is understood as *empirical* and *technological* success, exemplified by correct predictions, including those regarding manipulations or interventions on nature. Since Laudan argues *against* the Grand abduction of realism (as found in Putnam’s no-miracles), success cannot be construed too demanding: science ‘as a whole’ should come out successful. Reference is understood as a success term: a theory may, in some sense, ‘refer to’ non-existent, but let us interpret reference as *genuine* or successful reference.

There are various more or less demanding theories of reference on offer—for instance, by description or by pointing, baptizing and communicating—but as long as reference does not entail full truth, reference is insufficient for success. Examples are generated *ad libitum*: take a theory of which you believe that the terms refer, mess up some central claims and you may arrive at a very unsuccessful, referring theory. Bluntly put, “‘George Bush is fat, blonde, eloquent and atheistic’ refers to Bush all right, but would not be much good at predicting Bush-phenomena.” (Musgrave 2007) Real examples of referring theories that temporarily lacked success offered by Laudan in-

clude the Proutian theory of atoms and the Wegenerian (proto-)theory of plate tectonics. Now, it may be objected that such theories do not refer at all if reference is properly construed. Yet, if reference demands that all or most claims about the referent are true, in which case some of Laudan's examples may be crossed out as non-referring, this immediately gives us *less reason* to suppose that our current theories refer (because the realist cannot be burdened with believing in the full truth of current science). If there is some leeway between reference and truth, it is possible to construct counterexamples similar to the above ones. Hence, the inference of the antecedent ( $R \rightarrow S$ ,  $S \rightsquigarrow R$ ) is already risky, but reference does not even imply success.

It may be the other way around: success implies reference. Alas, this is exactly (part of) what the realist was trying to establish, and he cannot reliably abduct reference from success. To combat the alleged relation between success and reference further, Laudan (p33) presents a list of once successful theories that, by current lights, do not refer. For instance, Fresnel's ether theory of light predicted the occurrence of a bright spot at the centre of the shadow of a solid disc. This was a surprising novel phenomenon, but his luminiferous ether is nowadays considered a fiction. This list further includes Hippocrates's humors, phlogiston, the caloric theory of heat, and is claimed to be extendable *ad nauseam*.

Some of these examples may be rejected as *immature* science. This would mean that success does not imply maturity. One suggestion would be to restrict the explanans (the theories that are supposed to explain the success) to those theories that enjoy empirical success for a wide variety of phenomena. This may dispel some of Laudan's examples, although the restriction might be more naturally applied to the explanandum (the success to be explained), such that success is not explained by the approximate truth of mature theories, but maturity (success over a varied scope of phenomena) is explained by approximate or partial truth.

It is also open to realism, Laudan concedes, to claim that only *specific components* of successful theories refer—and this has indeed proved a popular strategy—but then the notion of 'approximate truth' has to be qualified accordingly. Truth *simpliciter* implies that all terms of a theory refer. Since some of the once successful theories on Laudan's list are nowadays considered to have non-referring terms, success does not imply full truth.

Putnam already qualified the truth-claim; no realist believes that present-day successful science is true *full stop*. What he wants to establish is that (we have good reasons to believe that) present-day successful science is near the truth or approximately true. The assertion that this truth-likeness is the best or only explanation of current success should compel us to infer this truth-likeness. But while we all agree that truth implies success, this is not so straightforward with approximate truth. As long as this concept is not clarified, it remains an open question whether it would explain success, so Laudan claims (p32). As it stands, the explanation by approximate truth

makes current success no less miraculous.

Laudan's list leaves open some loopholes and has prompted various realist refinements of explanans (truth), explanandum (success) and conclusion (realism). Also, it elicited the intuition that we are nowadays in no fundamentally different epistemic position than our scientific ancestors who abducted approximate truth from success, and we are equally overly optimistic. This is the pessimistic induction over the history of scientific theory change: assuming that past failures are representative, our best present-day theories are probably false. Let me stress that the 'inductive inference' is not rationally compelling; it is simply a rival intuition. The no-miracles abduction is non-monotonic. It proceeds from observed success, but if we add to the premises the observation of past failures, the inference to present-day approximate truth may be less appealing.<sup>1</sup>

### 2.2.3 Stance-Off

There is one general point that can be made without taking heed of the subsequent refinements of the structure and components of abduction: whenever the conclusion goes beyond the actual observable, it is, in the context of the realism debate, a *petitio principii*.<sup>2</sup> The empiricist condemns the adoption of beliefs about the supra-empirical as reckless epistemic behaviour. Against realism, this argument from recklessness begs the question, since realism does not share the empiricist epistemic outlook, and *vice versa*: the Grand abduction from the success of science to its approximate adequacy to the supra-empirical simply mirrors those more restricted realist abductions towards the 'truth-likeness' of specific theories, which are equally rejected by empiricism for going beyond the permissible evidence.

CE holds that explanation has to end somewhere; on pains of an infinite explanatory regress, we have to be satisfied with accepting *brute facts* at some point. Just saying that explanation has to come to an end somewhere, does not settle at which point we should stop explaining. However, the empiricist has a relatively definite and, from his stance non-arbitrary and well-motivated point beyond which explanations should not venture, i.e. the empirical. The realist claims that we hit bedrock with asserting the truth of our theories; the empiricist claims we hit epistemic bedrock with empirical adequacy. This does not mean that all should be explained by asserting science's empirical adequacy, but that explanations that *transcend* empiri-

---

<sup>1</sup>Worrall (1989) synthesized these two competing intuitions and made the case for *structural* realism, in a nutshell: the structural content, as opposed to the substance-ontology, of once successful theories has typically survived theory change and is sufficient to explain the success. For instance, although the ether associated with Fresnel's theory of light did not survive, his *equations* on reflection and refraction were retained in Maxwell's and, arguably, onwards in quantum electrodynamics.

<sup>2</sup>This is already pointed out by Laudan (1981), see also Fine, (1984), Musgrave (1988) and Lipton (2004, c11).

cal adequacy are epistemically unacceptable to the empiricist, regardless of their alleged explanatory virtuousness. When the only explanations on offer explain *by* making supra-empirical claims, the *demand for explanation* is rejected by empiricism. When doxastically digging deeper is epistemically reckless, we are satisfied with brute facts.

This rejection of the demand for explanation in specific cases is nicely illustrated by Van Fraassen's (1985) two-staged reaction to an argument by Hacking (1985). Hacking starts his article "Do We See through a Microscope?" (Hacking: yes, Van Fraassen: qualified no) with an elaborate and engaging exposition of the development of the microscope, the skill it takes to interpret, the struggle to combat the chief aberrations encountered, the invention of new methods to manipulate the relevant interactions, and so forth—truly a marvellous story of technological success. One especially staggering example of this success is this: if we have two microscopes that make use of different physical techniques and we put the same, perhaps slightly differently manipulated, observable slide under both instruments, then we get a very similar image. "It would be a preposterous coincidence if, time and again, two completely different physical processes produced *identical* visual configurations which were, however, artefacts of the physical processes rather than real structures in the cell." (Hacking 1985, p145, my emphasis)

There is no question whether these instruments repeatedly and reliably produce similar results; they do. Yet, what may we infer from this? The *first* stage of the empiricist reaction is to look for an explanation that need not appeal to adequacy to the unobservable. Note that the images will be similar, perhaps very much alike, but not identical. The instruments are fine-tuned while putting slides under it; we adjust a few bolts and nuts, interpret certain aspects of the resulting images as artefacts and others, plausibly based on interplay with the coordination of the other instrument, as *representative of* the unobservable content of the slide (i.e. the alleged real structures in the cell). In sum: the similar features that we perceive with both instruments are *selected for*. No wonder that these are persistent. (Van Fraassen 1985, p298) Admittedly, this alternative explanation is counterintuitive. The skilled user has learned to distinguish artefact from non-artefact and may be convinced he looks *through* the instrument. Yet, the alleged interaction with the relevant unobservables is *mediated*: there is an extra inferential step needed to establish their existence.

Of course, the realist can press the point if he is not satisfied with such an explanation: but what, then, *causes* these images if it is not the real structure in the slide? At this point, the demand for explanation can only be rejected. (Note that the empiricist does not deny the existence of real, unobservable structure, she only denies our epistemic access to the unobservable; as far as science is concerned we may remain neutral.) The *second* stage is to expose the question as forcing us to explain by postulation: to believe in

supra-empirical truth.<sup>3</sup> These two stages are mirrored when explaining the success of science as a whole is at stake.

For the record: the empiricist holds experiment in high esteem. Insofar we can speak of an empiricist ‘tradition’ to be discerned from or imposed by our present-day perspective, it surely emphasizes experimental practice, arguably starting with Bacon’s call to take up the instruments and subdue nature to manipulation. Yet, we have to keep in mind a helpful conceptual distinction due to Van Fraassen (2001, p154–63; 2008, ch4): we can naively think of instruments as windows to the unobservable world, but we had better think of them as *engines* generating new phenomena. The instruments produce observables and these are the target for responsible scientific beliefs.

As such, experimental practice gives a great impulse to science, stimulating the creation of novel phenomena to which our theories can be compared, whereas these theories suggest, in turn, new ways of how they can be put to the test experimentally. This interplay catalyses the quest for empirical coverage. The images, pointers, bleeps and other indicators aiding measurement need to be saved and predicted themselves. Yet, we should be aware that the leap from the reliable prediction of these phenomena towards the supra-empirical adequacy of the predictively successful theories, indeed, involves an *extra inferential step*. This step is in the empiricist’s eyes a step too far.

It has been argued that CE *relies* on abduction, but arbitrarily restricts the conclusions to claims concerning the observable. The accusation is that CE needs abductions to the observable, but rejects abductions beyond, while this distinction lacks a principled basis. The structure of the inference is the same, so if Van Fraassen accepts the first form, he should accept the second—or so the *tu quoque* charge goes.<sup>4</sup> This accusation is wrongheaded in various

<sup>3</sup>Hacking (*Ibid.*, p146–8) also presents his ‘argument of the grid’ which is essentially the same argument. The example is used here for expository purposes, but I should note that the debate is on-going. In a reply to Teller, Van Fraassen has conceded that he is prepared to allow looking in the optic microscope as an acceptable case of observing what is put under it, as long as it is recognized that we do not *look through* an electron microscope. (2001, p162–3) I do not see why this concession is needed, although Van Fraassen’s point is retained as long as we recognize clear cases and counter cases of direct observation. Further see Alspector-Kelly (2004).

<sup>4</sup>See the exchange between Psillos (1996, p34; 1997) and Ladyman et al. (1997). Niiniluoto (1999, pS449) and Lipton (2004, p199–206) echo the misguided charge, which Lipton dubs the ‘same path, no divide argument’, but bases it on the misunderstanding that CE claims “that our inductive powers extend only to the limits of the observable” (p154). Van Fraassen’s initial example of abducting the company of a mouse from cheesy tell-tales in the kitchen (1980, p19–20) has caused some confusion, for it was not made sufficiently clear that this abduction to empirical adequacy was an example of an optional inference, not compelled by the canons of rationality. Interestingly, the same example is later described by Van Fraassen as a typical *modus ponens*: “I already thought all along that if there are such changes in the kitchen scene then it was visited by a mouse.” (2005, p124) Still, Van Fraassen accepts that there are many cases of inferential behaviour which are, at minimum, in accord with (some explication of) abduction. (2006, p133)

ways.

*First*, CE does not ‘rely’ on abduction if this is interpreted as mandatory rule-following; although science, obviously, involves claims that go beyond the evidence (ampliative claims), there are no ampliative *rules*. *Second*, and more important for present purposes: the empiricist *has* a principled and non-arbitrary distinction: empirical vulnerability. Therefore, *even if* he were committed to abduction, he could consistently abduce the empirical adequacy of the explanation and retort that the realist illegitimately and artificially collapses the distinction between abducing empirical adequacy and recklessly abducing truth by pretending that we ‘normally’ abduce truth. Yet, as said, CE does not depend on abductions to the observable: explanatory power is deemed to be a pragmatic virtue and the relation of support between ampliative beliefs and evidential—including explanatory—beliefs is up to the believer.

The no-miracles intuition, thus, relies on a mode of reasoning that is rejected by the empiricist. Empiricists and realists take a different stance on science and although Putnam presents realism as a *scientific* explanation, this is not accepted by CE. Realism is no scientific theory that stands on ‘neutral’ ground. As intimated in the previous Chapter, it is hard to adjudicate between the rival views of science, for they rely on a prior differentiation between epistemic and pragmatic virtues. Realism deems explanatory success an epistemic virtue, CE does not. As a scientific theory, realism predicts that accepted theories remain empirically and explanatorily successful. CE holds that full acceptance of a theory involves the expectation that the theory remains empirically successful, but the pragmatic commitment likewise involves the expectation that the programme, including its explanatory virtues, will prove to be fruitful in the future—still, explanatory success is not a symptom of truth.<sup>5</sup>

Hence, both parties are dug in well and comfortably in their epistemic trenches; it is time for a stance-off. The realist needs to argue for the alleged epistemic import of explanatory virtuousness without already tacitly assuming that it has. In the remainder, I shall avoid reference to evidential invulnerability, because this concept has no common-sensical interpreta-

---

<sup>5</sup>Although Psillos initially presented his no-miracles intuition as leading to an *empirical claim* (1999, p80), he has meanwhile come to agree that the dispute is not a ‘neutral’ scientific issue: “The [no-miracles] is not an argument for scientific realism; that is, it’s not an argument for the truth of realism. [...] Scientific realism is not a theory; it’s a framework which makes possible certain ways of viewing the world. [...] For the [no-miracles argument] to work at all, it is presupposed that explanation—and in particular explanation by postulation—matters and that scientific theories should be assessed and evaluated on explanatory grounds. Hence, the no miracles argument works within the *realist* framework; it’s not an argument *for* it.” (Psillos 2011b, p33) Also see Ghins (2001), Worrall (2011) and Van Fraassen (2006, p138–41). The empiricist stance, hence, has a new *package-deal* rival in what Psillos has described as ‘the realist framework’, which puts explanation center stage. (Psillos 2011a, specifically p311–2)

tion. Instead, I show that realism is unable to make sense of its own alleged source of information: the evidential import of explanatory loveliness remains a complete and utter mystery.

### 2.2.4 Structure of the Inference

Psillos (1999, p81–90), following Boyd (1983), has attempted to infer the overall reliability of abduction based on the success of specific scientific abductions. At first sight, this inference is blatantly circular, because the overall reliability is proposed as the best explanation of specific success, thereby presupposing abduction to be evidential. On second thought, it is still blatantly circular, although Psillos tries to avoid the sting by appealing to a difference between ‘vicious’ *premise* circularity and ‘innocuous’ *rule* circularity due to Braithwaite.<sup>6</sup> Basically, instead of adopting the explicit premise that we may infer (the probable approximate truth of) the best explanation, this premise is left out, while the conclusion is still inferred, thereby relegating the explicit premise to an implicit rule. “If the rule of inference *is* reliable (this being an objective property of the rule [in an externalist sense]) then, given true premises, the conclusion will also be true (or better, likely to be true—if the rule is ampliative).” (Psillos 1999, p83)

Suppose I want to infer  $p$ , but  $p$  is under discussion, so I cannot add  $p$  as a premise; then I omit  $p$  and adopt the rule that I may infer  $p$ , like an introduction rule of the type: introduce  $p \vee \neg p$  *ad libitum*. (Cf. Worrall 2011, p20) If the rule to infer  $p$  is, as a matter of fact, reliable, then this inference is supposedly not viciously circular. Now, in a context of debate,  $p$  is apparently under discussion: this means that my opponent does not accept  $p$ , nor will he accept that inferring  $p$  is, as a matter of fact, a reliable rule. Therefore, I do not see the upshot of making the validity of inferences an external feature. Everyone agrees that *if* abduction is an objectively reliable rule, meaning that it preserves truth or confers likeliness, *then* abduction is reliable, et cetera. Yet, this does not help the initial discussion, certainly not since “the correctness of the conclusion depends [...] not on *having any reasons* to think that the rule is reliable.” (Psillos 1999, p84)<sup>7</sup>

<sup>6</sup>Or at least, rule circularity is not necessarily of the vicious type. Concerning the viciousness of premise circularity, Psillos has wavered a bit: “[T]he argument type ‘ $p$ , therefore  $p$ ’; should not be deemed viciously circular since, I take it, it purports to show that every sentence is a logical consequence of itself. But it would be viciously circular *were it meant to show* that  $p$  is true.” (1999, p82, my emphasis) This suggests that the viciousness of premise circular arguments depends on the dialogical context, but compare: “Premise-circularity [...] is *always and everywhere* vicious! It cannot possibly have any epistemic force for someone who does not already accept the conclusion.” (2011b, p25, my emphasis)

<sup>7</sup>Subsequently, Psillos concludes that realists had better revert to externalism, because their abductive inferences would otherwise be without justification! (1999, p85)—but this is withdrawn in his (2011, p26).

At best, the premise/rule-move shows that *if* you already accept abduction (to supra-empirical adequacy), *then* it works self-affirming, since it can be applied to reason from specific cases to overall reliability and back again. This is, of course, no independent argument for the epistemic import of explanatory loveliness, nor is it meant to be. In a later paper, Psillos (2011b, p23–4) breaks up the argument in two parts:

- A1. Scientific methodology (is theory-laden and) leads to empirical success.
- A2. This success is best explained by the approximate truth of the background theories.

B1. The scientific background theories are approximately true.

B2. These theories are abduced.

∴ Therefore, abduction is reliable.

To show that rule circularity is *as vicious as* premise circularity, let us suppose that we do not accept the concluded reliability of abduction (or that abduction is objectively not reliable if you will). Then, premise B1 is unacceptable. Where does B1 come from? A2 states that empirical success is best explained by approximate truth, but this purported explanation suddenly pops up in B1 as if it were an undisputed fact. (There is the implicit rule at work, I take it.) This ‘inference from the void’ is only permissible if the reliability of abduction is already presumed. On a side note, there may be an implicit abduction in the B-argument: it transforms a perceived correlation of ‘being abduced’ and ‘being true’ into the assertion that abduction, somehow, selects for truth. At minimum, it needs to be shown that false theories will (typically) not be abduced—in the light of Laudan’s list, this is highly contentious. *Even if* it could be established that *only* approximately true theories are abduced, Psillos might need a meta-abduction to conclude that this is no coincidence.

As it stands, the formulation is terribly vague. Intuitively, the ‘general methodology’, insofar this concept even makes sense, leads to both failures and successes, and success can be generated by false theories. For now, it suffices to drive home the point that the no-miracles intuition relies on a prior acceptance of abduction—whether as an explicit premise or hidden in the inference rule. To be fair, Psillos has recently conceded as much: “It’s not as if [no-miracles] should persuade a committed opponent of realism to change sides. But it can explain to all those who employ [abduction], in virtue of what it is reliable; and it can possibly sway all those who are neutral on this issue.” (2011b, p26)

## 2.3 Refining Abduction

### 2.3.1 Explanans: Particularise

Scientific success is supposed to be explained by the approximate truth of scientific theories. Putnam's reference to 'mature science' and Psillos's 'background theories' are in need of explication. The approximate truth of science 'as a whole' or the global reliability of the 'scientific method' are too vague to explain anything. Concerning the latter: a glance at Laudan's list teaches us that 'the scientific method' is no guarantee to success, because it also leads to failure.

Also, there is another, not necessarily competing explanation of the success of 'science as a whole' as proposed by Van Fraassen: "[A]ny scientific theory is born into a life of fierce competition, a jungle red in tooth and claw. Only the successful theories survive—the ones which *in fact* latched on to actual regularities in nature." (1980, p40) This explanation is often disqualified, for it cannot account for the success of *specific* theories. (*Ibid.*, fn34; Lipton 2004, p194) It is like explaining why all members of the sorority Oasis are dark-haired by claiming that only dark-haired members are allowed: it does not explain why Barbara, who is a member, has dark hair. In the next Section, I show that this is a false analogy, but for now we note that the Darwinian explanation removes the wonder of 'success at large'. The appeal of the no-miracles intuition is better exploited by focussing on *specific* theories. Instead of a wholesale abduction to realism, the intuition should be applied case by case. This is the first move towards particularisation. (If the approximate truth of only one theory should be abduced, I see no principled reason for barring the evidential import of explanatory considerations on a general level: I am prepared to convert to realism if only one abduction is compelling, which I deny.)

Success, then, is supposed to be explained by the approximate truth of specific theories. Yet, what does approximate truth mean? Various general measures have been proposed, but there is, at present, no agreed upon solution. (Psillos 1999, ch11) More promising proposals explain success by the full truth of specific *components* of a successful theory. For instance, the micro-constituents that the theory centrally postulates exist (Cartwright 1983, ch5; Hacking 1984), the truth-claim is restricted to capacities (Cartwright 1999), or to the structures involved (Worrall 1989).

This second move towards particularisation also makes sense because success is not explained by components that do not contribute to this success: the explanatorily *idle* components. Suppose we have a successful theory  $T$  and add to it the claim  $C$  that cherubim have wings.  $C$  might be such that it does not add any empirical coverage to  $T$  and does nothing qua *explaining*  $T$ 's empirical success. By inferring only those components of a theory that are 'needed' to explain the success, the abductor is less vulnerable to the

counterexamples on Laudan’s list. (Cf. McMullin 1984; Psillos 1999, ch5)

The precise explication of how a component can be ‘essentially responsible for success’ does not concern us here, nor does the nature of the various components proposed. We take the explains to be the truth of *<specified components>* of the theory and assume that this particularizing strategy makes the abduction less vulnerable to the items on Laudan’s list, although these items still show that abduction does not lead invariably to truth.

### 2.3.2 Explanandum: Success

We tend to be more impressed by the correct prediction of novel *types* of phenomena. The observation of the next black crow might be explained by the truth of the hypothesis that all crows are black, but if a theory correctly predicts a novel, unknown type of phenomenon, such as the bright spot in the shadow of the disc, we are very impressed. This has led several authors to consider restricting the explanandum to the correct prediction of novel types of phenomena or at least emphasizing that this success is inexplicable if not for properly particularised truth (Leplin 1982; McMullin 1984; Musgrave 1988; Worrall 1989; Psillos 1999, ch5; 2009, ch3; Lipton 2004, ch10).

It is suggested incorrectly that successful prediction is, *in general*, inexplicable on Van Fraassen’s Darwinian alternative: “The real miracle is that theories we judge to be well supported go on to make successful predictions. The selection mechanism does not explain this, since it does not explain why our best supported theories are not refuted in their next application.” (Lipton 2004, p194) However, this charge rests on a misreading of Van Fraassen’s alternative (*vide supra*). The Darwinian explanation holds that surviving theories may have latched on to empirical regularities, where this ‘latching on’ includes the ‘higher theory’ giving rise to correct generalizations. If a theory describes an empirical regularity, it is no wonder that it makes successful predictions and that it is ‘not refuted in its next application’.

Yet, when novel *types* of phenomena are concerned, the Darwinian alternative fails: the theory cannot have been selected for its latching on to an undocumented regularity. If a theory makes a *surprising*, correct prediction, this strengthens the intuition that the success can only be explained by the truth of the components responsible.<sup>8</sup>

---

<sup>8</sup>Novel success is also explained by the theory being empirically adequate—beyond which the empiricist will reject the demand for explanation. Some have called empirical adequacy a bad explanation, but this disqualification flows from realist presuppositions. I will not extensively compare the explanations by truth and empirical adequacy, for realism is supposed to stand on its own feet, but let me note that much criticism on the explanatory status of empirical adequacy is *off target*: “[A]ll the proposed non-realist explanations of [...] empirical success seem to fail. In particular, the attempt to use the empirical adequacy of a theory T (i.e. the fact that all the empirical consequences of T are true) as the explanation of its empirical success [...] leads to a trivial non-explanatory tautology, if empirical success means the ability of T to yield true empirical consequences”

Yet, there is something off with the high esteem we have for the successful prediction of phenomena that happened to be undocumented. Suppose that a scientist has inferred a successful surprising prediction from his theory, but it turns out that this phenomenon was already documented unbeknownst to him. Does this rob the theory of its impressive feat? Intuitively, one might say no: it is equally impressive as long as the scientist was unaware of the documentation. However, this has the counterintuitive consequence of transforming ignorance into an epistemic bliss. Intuitively, it should not matter for a theory's explanatory power whether a phenomenon was documented a day earlier or later and whether the scientist had heard of it before.

Let us play it safe and not burden the realist with restricting his explananda to the prediction of novelties. This may also narrow down too much the scope of theories of which the particularised truth can be abducted. In the light of Laudan's list, it seems most prudent to explicate success as empirical success with respect to a wide variety of phenomena, with the tacit understanding that especially surprising success calls for an explanation. This is summarized as 'empirical success'.

### 2.3.3 Conclusion: Likelihood

We have given somewhat more content to the premises of the no-miracles argument. Let us keep the abductive rule, which allows us to infer the conclusion, implicit: this does not matter for the validity of the inference and we keep in mind that making it implicit does not make it beyond scrutiny. Then we have the following genus:

1. Specific scientific theories are empirically successful.
  2. This success is *best* explained by <specified components> of the theory being true.
- ∴ Therefore, ?

Two questions to the friend of inference to the best explanation emerge: what is the inference to be made, and what is *the best* explanation? Without delving into the different accounts of explanation, we ask: should we infer the best (the only) explanation on offer, imaginable, possible or...? Assuming

---

(Niiniluoto 1999, pS448) The empirical success that is to be explained is typically *not* the *general ability* to yield true empirical consequences, but the *success observed so far*, which is implied by empirical adequacy, just as it is implied by truth: only entailment and no tautologies there. Niiniluoto says that the ability to yield true empirical consequences needs to be explained, but this phrasing is suggestive: we could just as well say that 'the ability' to yield true consequences needs to be explained. Being true is the same as having only true consequences, being empirically adequate is roughly the same as having true empirical consequences. As intimated, the realist and the empiricist disagree on where we reach explanatory 'bedrock'.

that the best *possible* explanation is *ipso facto* true, this does not help us in determining whether explanatory considerations are epistemically virtuous. Let us momentarily grant that the explanations of present-day science (including explanation by empirical adequacy) are superior to any alternative speculations that we imagine *ad hoc*, in which case the best explanation *on offer* is among the scientifically entertained explanations.

The realist may want to infer that the best explanation is true. And since we replaced the opaque ‘approximate truth’ by the full truth of specific components, he might be tempted to simply infer the unqualified truth of the claim to the unobservable adequacy of those components. But, let us take a step back. The second premise is *comparative*: the explanation is supposed to be superior to its rivals. Even if it is the only explanation we happen to have concocted, it is still comparative: its monopoly is due to a comparative lack of rivals. Yet, the conclusion to truth is *absolute*.

Van Fraassen has put forward an argument against abduction, known as *the argument from the bad lot*. (1989, p142–50; 2006, p132–41) Our set of explanations, perhaps with only one member, is a historically situated, contingent collection. These explanations might be compared to each other in terms of how well they are ‘supported by’ the evidence in order to decide which one is ‘the best’, but the abductive *recipe* requires that the best one is true full stop *and* that this true one is among the available ones. What to conclude when all available explanations are false—when the best of the available explanations is merely the best of a bad lot? If we came up with only one explanation, the abductive recipe prescribes that this explanation is *ipso facto* true—by fiat or lack of imagination, one might add.

Van Fraassen anticipates three realist replies: (A) privilege, (B) *force majeure* and (C) retrenchment. (A) *Privilege* amounts to the claim that scientists are, somehow, *predisposed* to hit on the right hypotheses, in which case we have good reasons to suppose that the true hypothesis is among the available ones. This purported innate *tendency to truth* could appeal to e.g. rational insight, God or evolution. The non-evolutionary alternatives either posit yet another source of knowledge, which falls beyond the scope of this Chapter, or, if our ‘explanatory insight’ is supposed to coincide with the source of extra-empirical information, such that the deeper source *only* expresses itself by way of explanatory insight, this is a matter of faith, which cannot be argued for independently. This is not the place to delve into evolutionary epistemology, but note that it is difficult to reconcile this appeal with Laudan’s observation of past failures.

(B) *Force majeure* is the reply that we must choose between our current explanations; good or bad, we have no others. Alternatively: the demand for explanation cannot be rejected. This strategy is arguably exemplified by Psillos’s “Choosing the Realist Framework” (2011a). The central claim is this: “[S]cience aims at explanation, which *indispensably* involves positing microscopic constituents of macroscopic things.” (p302) The positing of

unobservables is, according to Psillos, indispensable to achieve ‘causal and nomological coherence’ which is needed to understand and predict the observable realm. (p309) Mundanely put, we must understand the table in terms of its alleged unobservable building blocks (and believe in their existence) to make sense of what happens when we and other objects interact with it, or to explain why it is stable and does not evaporate. I was tempted, for a moment, to redub this line of argument a *farce majeure*, for it is just a rehearsal of the realist tenet, but it turns out to be, as the title of the article suggests, a force *optionnel*: “[i]ndispensability arguments work only relative to accepting certain aims.” (p310)

Obviously, CE has an alternative portrayal of science in mind, but we should evaluate the merits of the realist framework on neutral terms. The pivotal question, then, is why we should accept the framework and the resulting esteem for explanation. Psillos offers, in good conscience, no ‘frame-transcendent’ answer to this question, except for an offhand allusion to Grover Maxwell, who concurs that any science-oriented framework must put explanation centre stage. (p311)

The general suggestion that, in science, we must value explanation above anything else—i.e. that we should adopt the realist framework—has smaller scale analogues. Recently, pastafarianism, a religion that originated in 2005, discovered a correlation between the global rise of temperature and the decline of the number of pirates in the world. In their eyes, the correlation is highly significant, for they attribute the cause of global warming to their deity, who happens to be fond of fun-loving buccaneers and aims to punish mankind for their decline. Admittedly, this example is artificial, but the point will be clear enough: we cannot be *forced* to explain by postulation, not even in the absence of any alternative explanations. The pastafarian explanation is no serious contender, but also ‘serious options’ may be rejected if they are *not good enough*.

Psillos’s claim is conditional: *if* we accept the demand for explanation, *then* we are committed to science’s explanatory posits. Now, if the framework forces us to believe all ‘accepted ontology’ in one fell swoop, then the framework is unacceptable: on a smaller scale, in specific cases, explanation is not sacred, even if it allows us to postulate otherwise inexplicable connections. It is *not* the case that “the choice is constrained.” (p311) If this demand is qualified, as it should be, such that we can differentiate between epistemically un/acceptable explanations, then explanatory posits are, in general, apparently *not indispensable*.

Thus, our evaluation of the indispensability argument depends on what is meant by ‘the explanatory posits’ or ‘the accepted ontology’. Once more: if this encompasses all posits of the ‘best’ explanations that happen to be on offer, this is a reason to reject the realist framework: those explanations should be examined case-by-case and are not indispensable *sui generis*. This articulation is sometimes suggested by Psillos, for instance: “If there was only

one potential explanation, it would be folly not to accept it.” (2009, p198) On the other hand, if the explanations are examined case-by-case and may be rejected if they are not good enough, then their supposed indispensability is rendered questionable. This reading is suggested when Psillos qualifies the indispensability: “which entities we have reason to believe are real is a function of the degree of confirmation of scientific theories.” (2011a, p311) From other writings (*vide infra*), it appears that Psillos actually opts for this route of ‘selective indispensability’, thereby rendering moot the force majeure as a Grand *inference* by indispensability.

(C) We cannot be forced to infer unconditionally that the best of the explanations on offer is true. This leads to the third of realist replies, which Van Fraassen calls *retrenchment*. This reply rejects that we should always infer the truth of the best explanation. Instead, it claims that the comparative explanatory virtuousness mirrors or should mirror a comparative likeliness, i.e. explanatory virtuousness has *evidential import*. Psillos already implied the dispensability of epistemic commitment to explanatory posits, when he rendered the abductive inference conditional: “in so far as the best explanation is likely enough to allow an inference to be made.” (1996, p32)<sup>9</sup> When this additional constraint is added, we are no longer supposed to infer ‘blindly’ the truth of the comparatively best explanation.

Yet, we must be careful in formulating what is meant by an explanation being ‘good enough’. If the explanation should be *sufficiently likely*, and likeliness is spelled out in terms of subjective, rational credences, then this might well put the cart before the horse. If we already believed that some theory is relatively likely—or sufficiently likely to infer—explanatory considerations are *redundant*. The point of the abductor’s exercise is to convince us that explanatory virtuousness is itself a *symptom of truth*.

## 2.4 Loveliness and Likelihood

### 2.4.1 Symptom of Truth

Lipton (2004, ch3) is astutely aware of the danger to get ahead of the conclusion and unpacks the ‘inference to the best explanation’-slogan as the less catchy, but more accurate: inference to the *loveliest* of competing potential explanations under consideration, if sufficiently lovely. (The ‘potential’ signals the view that every genuine explanation is *ipso facto* true, cf. *Ibid.*, p57, and ‘competing’ that explanations, especially the causal ones, may point towards different salient factors without being actual *competitors* for the inference, viz. they can be complementary, cf. p63.) Lipton gives no definite

<sup>9</sup>The qualification that the explanation should be good enough is repeated in Psillos’s (1999, p79) and encountered frequently, for instance in Kvanvig (1994, p328) and Cartwright (1983, p87)—who, incidentally, also confuses likeliness and loveliness (*Ibid.*, p6). This conflation is corrected in Psillos (2007a, p442–3).

threshold of sufficient loveliness (explanatory virtuousness), nor am I aware of any proposals to this effect, nor can I imagine what an answer would look like (.8 lovely?). Without a definite threshold, abduction is transformed from a recipe for inferring full beliefs into a recipe for altering *degrees of belief*.

In general terms, then, we can say: the lovelier, the likelier—this can either be interpreted descriptively or prescriptively. If there are several alternative explanations, we can rank them in terms of loveliness and claim that the loveliest is the likeliest. If there is only one explanation, and it is very lovely, we can conclude that it is very likely. When loveliness reaches a certain (as of yet unspecified) threshold, we may have sufficient reason to believe it is true. If (some components of) our current best scientific theories reach this threshold, we convert to (a restricted type of) realism. This general claim, the lovelier the likelier, permeates the various proposals and intuitions appealing to explanatory considerations. It is only via this route that Putnam’s intuition may be substantiated after all.

Lipton wishes to leave ‘loveliness’ a blank, to the effect that we can plug in our own favourite account of explanation. He does say it should have something to do with providing us *understanding*. With respect to the meaning of likeliness, he is not very clear, although it has something to do with ‘speaking of truth’, ‘warrant’ or ‘probability’. (p59) “Even without the evidence that favored special relativity, the production of the theory *probably* made Newtonian mechanics *less likely, but probably not less lovely*.” (*Ibid.*, p60, my emphasis) Let us regiment vocabulary and let ‘plausible’ refer to our personal judgments of a theory along the *informal* lines of, ‘that theory may very well be true.’ Then, at least, we can say that loveliness *should* inform our judgments of plausibility.

To say that something is likely, is to say that it has a high probability, but this is interpreted in various ways. One idea is that probability is something in the world, independent of our conception of it. Such objective interpretations include *frequentism*, according to which the probability of  $e$  refers to the proportion of  $e$ -type outcomes relative to the total set of outcomes of some chance setup (reference class). This total set of outcomes may consist of all *actual* outcomes (past, present, future) or the outcomes of all *hypothetical* trials, rendering the subspecies actual and hypothetical frequentism. The set of actual outcomes is typically finite, while the set of hypothetical outcomes is usually thought of as being infinite, in which case the probability of  $e$  is its limiting relative frequency. Another objective interpretation is the *propensity* account, according to which probabilities are somehow physically realized, causally efficacious dispositions or tendencies of some setup to produce certain outcomes.

Thus, the objective probability (chance) of some coin (setup) toss (trial) landing heads (outcome) may be interpreted as (i) the proportion of heads in the set of all actual (past, present and future) tosses of this coin; (ii) the proportion of heads in the limit of the infinite set of hypothetical tosses of

this coin; (iii) the tendency of this coin to produce heads.

I do not know how to make sense of ‘the lovelier, the likelier’ on the propensity account, but along frequentist lines we could make the claim that in the set of lovely theories a higher proportion of them is true than among the unlovely theories, so that a lovely theory has an objectively higher *chance* of being true than an unlovely theory. This would make the case that our judgments of plausibility *should* be informed by judgments of loveliness. Here is Van Fraassen’s criticism on this construal:

I believe, and so do you, that there are many theories, perhaps never yet formulated but in accordance with all evidence so far, which explain at least as well as the best we have now. Since these theories can disagree in so many ways about statements that go beyond our evidence to date, it is clear that most of them by far must be false. [...] Hence it must seem very improbable to me that it is true. (1989, p146)

This criticism is *off target* (cf. Okasha 2000, p697–99). To claim that judgments of loveliness should inform our judgments of plausibility is not to claim that any (conceivable purported) explanation has a high chance of being true, but rather that lovelier theories are likelier than unlovely rivals. Van Fraassen’s construal heaps up *all* (purported) explanations—regardless of comparative loveliness—and makes the trivial case that most of them are false. Moreover, he suggests the application of the contentious ‘principle of indifference’ to the effect that we must treat all members of the probability space as equiprobable, i.e. assume a uniform credence distribution, in the absence of relevant evidence.<sup>10</sup> *First*, he does not accept this principle himself (1989, ch12) and, *second*, the explanationist (the friend of abduction) claims that he *has* evidence relevant to the probability distribution, namely loveliness.

When the reference classes are specified properly, the frequentist construal seems an acceptable way of spelling out the claim that loveliness is a symptom of truth. If it can be substantiated, the explanationist can make the case that loveliness should inform our judgments of plausibility by introducing a principle to connect our rational credences with our beliefs about the chances, such as Lewis’s principal principle. (Lewis [1980] 1986a)<sup>11</sup> Alas, it is difficult to substantiate, for we would need insight into the set of true theories. If we had this, explanatory considerations would be redundant.

Alternatively or additionally to the above interpretations of probability (this is debatable), there seems to be a use of ‘probability’ in everyday

---

<sup>10</sup>The principle that symmetrical evidence demands symmetrical expectations was dubbed ‘the principle of indifference’ by Keynes (1921, p42); he ascribes it to Bernoulli, who called it ‘the principle of non-sufficient reason’.

<sup>11</sup>For a refinement of the principle, see Lewis’s (1994). However, it is surprisingly difficult, if not impossible, to justify this principle in non-circular fashion, cf. Strevens (1999).

language that it not so easily captured by the objective interpretations (perhaps excluding the propensity account), exemplified by such a statement as, ‘Michaëlla Krajicek will probably not win the next Wimbledon.’ *Objective* interpretations (chance) are contrasted with the *subjective* interpretation of probability (credence, see Chapter 1, Section 1.3.1).<sup>12</sup> The explanationist could claim that our *rational* credences take into account explanatory considerations. In that case, our judgments of plausibility and judgments of likeliness collapse. The subjective interpretation of probability is commonly examined within the Bayesian framework.

### 2.4.2 The Bayesian Framework

While there are many variants of Bayesianism and many axes along which they can be distinguished (cf. Weisberg 2009), I take it that Bayesianism minimally extends ‘bare probabilism’ (the position that our credences should conform to the calculus in order to be rational) with the claim that we should update our credences in the light of new evidence according to Bayes’s theorem (Bayes 1763).

$$P(H|E) = P(E|H)P(H)/P(E) \quad (2.1)$$

While Reverent Bayes’s theorem is a theorem of the calculus, implied by the definition of conditional probability as  $P(H|E) = P(H \wedge E)/P(E)$ , *Bayesianism* is characterized by using the theorem to constrain the updating of our personal credences conditional on the evidence. Applied as such, it is called the rule of *conditionalization*. The probability  $P(H|E)$  is the probability of some hypothesis  $H$  conditional on the evidence  $E$ . When  $E$  is acquired, this is what we want to compute. The solution is the new or *posterior probability* of the hypothesis.  $P(E|H)$  is the probability of the evidence *given* the hypothesis, sometimes called *the likelihood* of  $E$ . This is  $E$ ’s probability according to  $H$ ; if the hypothesis entails the evidence, this is 1, if it is a statistical hypothesis, it is just the probability that  $H$  assigns to  $E$ .<sup>13</sup> The other two probabilities,  $P(H)$  and  $P(E)$ , are the *prior* probabilities, which are assessed with the help of our background knowledge. The prior probability of  $E$  is called  $E$ ’s *expectedness*. The prior probabilities are yesterday’s posteriors, but in the end, the original or *initial priors* are up to the believer.

Suppose we compute the hypothesis  $H$  that the die rolled on 6 given the evidence  $E$  that it rolled on an even number: the likelihood is  $P(\text{even} | 6)$

<sup>12</sup>These different uses of the term ‘probability’ were originally pulled apart by Carnap (1945).

<sup>13</sup>Since this updating rule is supposed to trade exclusively in *subjective* credences, we actually need a further principle to ‘force’ the believer to set the likelihood equal to  $E$ ’s chance according to  $H$ . See footnote 11. I will ignore this ‘probability coordination’ problem.

= 1,  $H$ 's prior is  $1/6$ , and the expectedness of an even number is  $1/2$ ; then  $P(H|E) = (1/6 / 1/2) = 1/3$ , which is the posterior probability of  $H$ .

Note that the posterior increases with the *unexpectedness* of the evidence, thereby incorporating the intuition that successful flabbergasting predictions give us more confidence that the theory is true. Yet, the lower the likelihood, the less confirmation, and if the likelihood is less than the expectedness, meaning that the evidence was more expected in the light of our background knowledge than according to  $H$ , then the evidence *disconfirms*  $H$ . Dis/confirmation is, here, interpreted qualitatively:  $E$  confirms  $H$  iff  $P(H|E) > P(H)$ . If the evidence was already established, i.e. the expectedness is 1, then  $E$  does not do anything for  $H$ , which is a rather harsh condemnation of accommodation; this feature is known as the problem of old evidence.<sup>14</sup> Finally, in contrast to 'Jeffrey conditionalization' (Jeffrey 1965; 2004, p57–9), orthodox conditionalization (2.1) models the acquisition of evidence as 'revelation', i.e. indubitable: there is no option to give the 'posterior of  $E$ ' a non-1 probability.

Recall that the norm of probabilistic coherence is defended by showing that if we translate your credences to hypothetical betting behaviour, and your distribution is incoherent, then you can be duped into a synchronic betting combination that is certain to lose you money, no matter how the events that you bet on turn out. This is known as a *Dutch book*. Synchronic incoherence leads to Dutch-bookability (Ramsey [1926] 1931; De Finetti [1937] 1992) and the converse holds as well. The rule of conditionalization can be defended by a comparable, so-called *Dutch strategy*, as presented in Teller (1973) and attributed to Lewis. Teller proves that if you commit to a rule for *changing* your credences in response to future information, then this had better be the conditionalization rule (2.1), for the adopting of any other rule makes you liable to a series of bets *throughout* time which result in guaranteed loss, i.e. makes you diachronically incoherent.

### 2.4.3 Objectifying Bayes

If the explanationist wants to incorporate explanatory considerations in the Bayesian framework, then he could complement conditionalization (2.1) with a rule to the effect that we are *rationally required* to judge 'the lovelier, the likelier'. The most natural way to spell this out is to let loveliness inform our prior probabilities of hypotheses. Two other options are: to let loveliness boost the posterior or the likelihood. The first of these alternatives, to let loveliness boost the posterior probability, is a non-starter. The idea is that the posterior probability of the hypothesis that best explains incoming evidence is given 'bonus points'. Alas, the adoption of this rule makes us vul-

<sup>14</sup>Some good books elaborating on the foundations and fruitfulness of the Bayesian framework are: Earman (1992), Fitelson (2001), Jeffrey (2004), Bovens and Hartmann (2004) and Howson and Urbach (2006).

nerable to Dutch strategies, as Van Fraassen has shown. (1989, p160–70)<sup>15</sup> Of course, the framework is not sacrosanct: if the abductor wishes to press his point, he could reject conditionalization (2.1) and the Bayesian framework. However, we explore the option of incorporation somewhat further as it is informative.

Instead of systemically altering the posteriors, one might let loveliness inform the likelihood. This is *prima facie* difficult to motivate. The likelihood is the probability of the evidence that is provided by the hypothesis under consideration. Why suppose that the better explanation renders the evidence more probable? Okasha (2000, p703) gives an example of two competing medical diagnoses,  $M$  and  $N$ , where the symptoms  $S$  are compatible with both, but more likely—indeed, “exactly what we would expect”—given  $M$ :  $P(S|M) > P(S|N)$ . I think this articulation has it backwards if it is intended as a demonstration of the normative force of loveliness. If  $M$  renders the evidence more likely, the reason for  $M$ 's higher posterior probability is *not* its explanatory loveliness. One could say that  $M$  is the better explanation *in virtue of* the relatively high probability it gives to the evidence. Then, the likelier is the lovelier, not the other way around. If loveliness is supposed to do some real work, then we should boost the likelihood based on explanatory virtuousness, in such a way that, irrespective of the initial likelihood, the lovelier explanation gets ‘bonus points’ on the likelihood.

A different way of pointing towards the same problem is to consider, with Psillos (2009, p196–7), cases where rival hypotheses initially provide the same likelihood, for instance: they both entail the evidence. If loveliness is supposed to enter the game via likelihoods, we should lower the likelihood of the less lovely hypothesis. But what can be the rationale to do this? A necessary condition to explain  $E$  is to entail  $E$  ( $H$  obviously cannot explain  $E$  if it entails not- $E$ ). If one of the competitors is less lovely, we might judge the theory itself to be less likely, but it is a peculiar move to reflect this intuition by changing the probability of the evidence conditional on the hypothesis. The likelihood reflects the probability of the evidence supposing that the hypothesis is true. If I suppose that some hypothesis that entails the evidence is true, then my credence of the evidence is—and should be—simply 1. The intuition behind abduction is that explanatory considerations should, minimally, help us in resolving evidential ties. If two theories render

---

<sup>15</sup>Interestingly, Van Fraassen has meanwhile taken the position that Dutch *strategies* are ineffective to expose irrationality. (1995b) Although this strategy was employed to defend his principle of reflection, Van Fraassen is still committed to reflection and holds that conditionalization is actually a special case of reflection and that *if* we are to follow a rule to govern our everyday credence kinematics, it had better be conditionalization (2.1). (1999) For an excellent case *against* the effectiveness of Dutch strategies, see Christensen (1991). Orthodox Bayesians who justify conditionalization (2.1) by Dutch strategies should, however, still be impressed by Van Fraassen's argument, since it is a valid instance of Teller's general scheme. The argument is further discussed in Leeds (1994), Kvanvig (1994), Niiniluoto (1999), Okasha (2000), Iranzo (2008), Weisberg (2008) and Psillos (2009, ch11).

the evidence equally likely, then their comparative loveliness may influence their posteriors, but it makes no sense to do this *by way of* altering the probabilities they assign to the evidence.

Considerations of loveliness are most naturally incorporated by altering the *prior* probabilities of the hypotheses. For the subjective Bayesian, these priors are ‘free’; they are the personal estimates of the likeliness of theories given one’s background information. The explanationist should complement the framework by introducing a rule that prescribes that we give bonus points to the priors of theories which are explanatorily virtuous, i.e. provide the more lovely explanations of the background information. We should be careful to distinguish between description and prescription. The *subjective* Bayesian permits any mechanism for choosing one’s personal priors: you may be guided by loveliness, but just as well by tossing a coin or reading tea leaves. It could be that we often, as a matter of fact, let loveliness inform our personal choices—this is Lipton’s (2004) *descriptive* thesis—but this is not sufficient to make abductions rationally *mandatory*. Whether we *do* does not settle whether we *should*. The abductor wishes to make explanatory considerations compelling, i.e. all are required to abduce unobservable adequacy, viz. realism.

Lipton proclaims to find himself in a Catch 22: he aims to show that our actual judgments of loveliness and judgments of likeliness often coincide, but on the other hand he wants to make the case that judgments of loveliness should be a real guide to judgments of likeliness. (2004, p124–5) If our intuitive judgments on these two features turn out to be such that they typically do not coincide, then his descriptive thesis is in trouble. But when a considerable descriptive overlap is established, it becomes unclear where the *epistemic* support comes from: is our posterior credence assignment based on loveliness or is it actually supported by likeliness whereas the judgments of loveliness are epistemically superfluous? In the latter case, Lipton’s normative thesis is in trouble. In the next Section, we examine cases where the intuitive judgments come apart. First, I emphasise that the mere *compatibility* of subjective Bayesianism and guidance by loveliness is insufficient for the explanationist.

Instead of claiming mere compatibility, Lipton claims that abduction can *complement* subjective Bayesianism. Where the Bayesian is silent, explanatory considerations come in as an ‘engine of inference’ or as a ‘central heuristic’. (2004, p107) For instance, if we have a hard time of estimating the likelihood (if the evidence is not entailed nor given a definite probability), assessments of loveliness may help us with assigning a likelihood-value. Also, explanatory considerations can guide our judgments of the priors if we have nothing else to go on. (p114–5) This route is also proposed by Okasha (2000). I said that his construal of likelihoods shows, at best, that judgments of loveliness are guided by judgments of likeliness and not the other way around. Yet, if loveliness is partly constituted by high likelihoods and

high priors, then we might employ explanatory considerations in assigning these values in the absence of other guiding information.

All this is fine for the subjective Bayesian—just as we are allowed to fill the gaps by other considerations. It does not show that we *should* let loveliness be a guide to likeliness, only that we *could*. We can make an analogous case for reading tea leaves or any other alleged source of evidence: it helps us in filling the gaps when we have otherwise no guidance and it can influence our posteriors by influencing likelihoods and priors. We can incorporate this in the framework, but why should we? Mere compatibility, or even filling the blanks in the absence of any better information, is insufficient. On the compatibilist construal, the normative force comes from the subjective framework. “[E]xplanationism can be understood in a way that respects Bayesian constraints and indeed in a way that *serves* a Bayesian approach to inference.” (Lipton 2004, p147, my emphasis) Normatively, this ‘service’ is to be interpreted literally, as it is “compatible with the denial of a strongly rule-governed inferential system.” (Lipton 2005, p356) Yet, this is of no help to realism. For the realist, explanatory considerations must be *normative*, and this means that they cannot be incorporated in a thoroughly *subjective* framework. Instead, she should overthrow subjectivism and constrain the choice of priors, i.e. she should opt for objective Bayesianism. (Cf. Psillos 2007a; 2009, ch11; Weisberg 2008)

## 2.5 Voltaire’s Objection

### 2.5.1 Explanation as Information

Why believe that we live in the loveliest of all possible worlds? This is, what Lipton calls, Voltaire’s ‘objection’—of course, after the author of *Candide*, who let Dr. Pangloss, one of the characters, say that the nose is designed to carry the spectacles (ridiculing Leibniz’s best of all possible worlds thesis). This is the pivotal question: what reason do we have to believe that explanatory considerations are *epistemically* relevant?

If we construe the epistemic aim as (a) believe all truths and (b) believe no falsehoods, then there is one obvious benefit, namely to believe an explanation is to believe more and thereby might aid us in arriving at (a). Yet, this can be said of any belief. The question should be: why is explanation *especially* relevant, aside from increasing the extent of the doxastic risk we take. In the previous Chapter, I construed ‘information’ as an endorsing term. Let us temporarily retract this and say that information is putative information; it provides doxastic options. Then the question becomes this: what is epistemically special about explanation aside from being information?

We have managed to steer clear of specific accounts of explanation and continue this course, but it is illuminating to briefly examine two sugges-

tions. Often-heard explanatory virtues include unifying power, simplicity, precision, comprehensibility and elegance. Take precision; as Barnes (1995, p260) argues, the prediction that I will meet a person of 2.12 meters tall tomorrow is more precise than the prediction that I will meet a person taller than 1.5 meters, yet it is also less likely (to be true). If precision is what makes for loveliness, it would be prudent to *avoid* loveliness—in the sense that avoiding precision promotes the epistemic aim of avoiding falsehoods. Barnes (1995) scrutinizes unifying power and the virtue of pointing towards a mechanism in similar fashion, as does Gijsbers (2011, ch3) with scope. However, both authors seem to focus purely on the epistemic aim of avoiding falsehoods, while ignoring the other aim: believing truths. Information-oriented activity should incorporate both desiderata.

The point I wish to make is this: sometimes likeliness and judgments of loveliness pull in opposite directions. This does not mean that we should infer the likelier story; if we want to avoid believing a falsehood at any cost, we might as well restrict our belief to tautologies. It simply means that believing some explanation usually involves believing additional information. And believing anything is always an epistemic risk. Yet we have no reason to suppose that the risk of believing more is somehow reduced if the information is explanatorily lovely. Indeed, in the case of precision or scope and the like, it is *not* the case that ‘the lovelier, the likelier.’

### 2.5.2 Internal Resistance

Psillos argues that we should doxastically rely on explanatory considerations, or on any way of reasoning for that matter, *unless* we have reasons not to: “a doxastic/inferential practice is innocent until proven guilty.” (2011b, p28) In Section 5.1 above, I have presented considerations to doubt the truth-conduciveness of explanatory considerations, and Laudan’s list shows, at least, that there have been abductions to theories that we nowadays consider to be false. Lipton claims that “insofar Bayesian inference avoids Voltaire’s objection, so can [abduction].” (2004, p148) With Lipton I agree that explanatory considerations are normative precisely to the extent that *subjective* Bayesian inferences are, i.e. not at all. Conditionalization is not ampliative. It presupposes that one’s initial credences are in place. Any ampliative recipes, then, are already reflected in the initial choice of priors and this choice is up to the believer. The updating rule is only intended to secure coherence: it is logical updating only.

Thus, we are allowed to take explanatory loveliness into account in our choice of priors, yet we have no positive *epistemic* reason to do so—indeed, neither do we have epistemic reasons not to. The considerations involved are just the standard epistemic considerations: believing more allows us to believe more truths, but also exposes us to a greater risk of believing falsehoods. Now, why are explanatory considerations *specifically* relevant

for our rational opinion management? I am not very fond of shifting the onus of justification, but an analogue is helpful.

Games are played by the rules in the manual. Discussion over the validity of certain *moves* remains *internal* to the ‘realm’ of the game. This means that the moves are not ‘frictionless’; they are very much restricted. Yet, they are not restricted by anything outside of the game world. I already alluded briefly to pastafarianism. Their deity is the Flying Spaghetti Monster (it should be capitalized, so I am told), which is unobservable by nature. This means that his followers can only detect His presence *indirectly*, by His ‘signs’. Now, if their religious framework is rejected, the signs of their deity might seem ridiculous, but, from within the framework, it is serious business and all kinds of rules can be devised to adjudicate between genuine detection and mere speculation that happens to be consistent with the ‘signs’. The Monster may have mysterious abilities, but if He is supposed to have caused something, it should preferably be explained in a way that we are able to grasp, perhaps by analogue to the happenings we see around us. Hence, the faith involved is not gratuitous and does not come for free.

In science, we (additionally) cope with a different type of resistance: the external, ‘worldly’ resistance. The intuition that our theories are only able to predict this external resistance in virtue of latching on to deeper, unobservable structures is well entrenched and tempting. Yet, the external resistance is *empirical*: first and foremost it settles which theories are inadequate and consequently the associated ontologies are overturned and revised. The only *additional* resistance that unrefuted ontologies are able to face is *internal*, i.e. of our own making; as Psillos remarks: “we cannot even start talking about the probability that there are electrons, or quarks or whatever unless we have already adopted the frame of theoretical entities.” (Psillos 2011a, p308)

It is hard to pull explanatory success apart from empirical success. In the end, (sensory) experience trumps theory, no matter how lovely. Loveliness may, as a matter of sociological fact, settle experiential ties, but how much pause should this give us? When two unrefuted theories have rival empirical consequences, we go on to put them to the test, *regardless* of their comparative loveliness. Suppose that we manage to latch scientifically on to a universal, empirical generalization. I take it that if we have formulated some unlovely theory *U* that ‘covers’ this generalization, then this theory will not be thrown away based on explanatory nausea. Rather, we keep it and, plausibly, make it lovelier.

From the mythological world views onwards, we have seen the Grand schemes of interpretation and associated standards of explanation rise and decline, while empirical coverage grows. Our theory *U* is made lovelier and is empirically successful or it is replaced by even lovelier *V* which retains the empirical success. What to make of the explanatory posits involved? They could be true; we could be lucky. They could be false. There is not

much to say about their adequacy to the unobservable and, if the theory is empirically adequate, its in/adequacy to the unobservable cannot be tested; it is subject to ‘internal criticism’ only.

Truth implies empirical adequacy and if a theory meets sufficient empirical resistance, we conclude its empirical inadequacy and falsity. To make the case for explanatory *evidence*, a comparable relation between truth and explanation should hold. But why suppose that truth implies explanatory loveliness? *False theories can be lovely* and there is no reason to suppose that true theories are lovely in virtue of their truth. Empirically successful theories or components of theories are retained and are ‘licked into shape’ according to the contemporary standards of internal criticism. Voltaire’s objection is aptly named.

To illustrate this, here is a model of the world (its adequacy depends on the interpretation):



This could be (mis)described in various ways:

- (i) There are two Q-circles with a symmetric edge, an R-square and an R-triangle.
- (ii) There is a triangle.
- (iii) There is no triangle.
- (iv) It has an edge iff it is both a circle and a Q, if not, it is R.
- (v) If it is a triangle, it is either R or Q, but if it is a Q, it is always a circle; no non-squares are P, whereas there is at least one R-square and if some shape has an R it is a unique shape *s*, i.e. no non-R *s*-shapes exist.

Some descriptions are more informative than others; they tell us more of the domain. Some descriptions are more simple or elegant than others. Along the lines of Lewis, one could say that we are fond of collections of correct claims which, as a system, strike the best balance between simplicity and strength, or elegance and informativeness.

There is no reason to suppose that the lovelier descriptions are likelier, e.g. (iii) is a very simple and false descriptive system. Yet, it could very well be that, once we have a true unlovely descriptive system, such as (v), we wish to make it more elegant. This would be a half-penny psychological account of the tendency to opt scientifically for loveliness. So here is a rival explanation of the perceived correlation between explanatory and empirical

success: after we have arrived at empirically successful descriptions, we make them lovelier.<sup>16</sup>

An alternative explanation of the perceived correlation between loveliness and empirical success is proposed by McAllister (1996, ch4–ch6): our judgments of loveliness are informed by empirical performance, so that the aesthetic ‘form’ of un/successful theories guides the future standard of loveliness. After an unlovely theory has enjoyed impressive empirical success, “in the longer term a correlation between scientists’ empirical and aesthetic evaluations tends to emerge.” (McAllister 1996, p65) This can be explained by the aesthetic standard, or scientific ‘fashions and styles’, shifting gradually with new generations of scientists. (*Ibid.*, p85) In both cases empirical success is what counts whereas we have no good reasons to suppose that loveliness has epistemic import. (Cf. *Ibid.*, ch6, especially p101) Loveliness is either built into the theory or our aesthetic standard is informed by it *ex post facto*.

We cannot dispel explanatory virtuousness as a source of genuine information, but I dare say that we (currently) have no more reason to take it thusly than to accept considerations flowing from any alternative, coherent framework of internal criticism.

### 2.5.3 Context of Construction

There is an important role for explanatory considerations in science. The suggestions by Lipton and Okasha to interpret them as a heuristics has its precursor in Peirce, who, at times, considered abduction a method ‘of hypothesis’ (Cf Niiniluoto 1999), i.e. a way of inventing new theories, instead of providing ‘warrant’, ‘justification’ or rational requirements. This was all along the proper way to conceive of loveliness. A preference for the lovely helps us to keep our hypotheses and ontologies tidy and can guide us pragmatically in theory choice, to the effect that the scientist commits to theories that he finds lovely and immerses himself, at least for a while. Loveliness is not an epistemic virtue over and above providing for (purported) information, but plays a fruitful and central pragmatic role.

In the context of CE, it is dangerous to talk of ‘the context of discovery’ or, in this case, ‘the context of construction’, since CE admits of no substantial objective context of *justification* of our ampliative scientific beliefs.

---

<sup>16</sup>For this reason, a methodological assumption about the universe or the empirical realm being simple cannot be motivated epistemically. We can describe it more or less simple: we can easily construct very unlovely but true descriptions about the above model. The universe makes our descriptive systems true or false and is *indifferent* to their comprehensibility or elegance. Some scientists are fond of beautiful or simple and useful theories, but this does not imply that simple theories are less likely to encounter anomalies—they do, repeatedly, and theories change: if recalcitrant evidence is acquired, science converts to alternative beautiful descriptions. So, there is no argument ‘from convention’ there, although simplicity is highly *useful*.

Actually, Van Fraassen sides with Hume's and Popper's criticism on inductive *rules*, while taking heed of Salmon's diagnosis that "science is inevitably inductive in matters of intellectual curiosity as well as practical prediction." (1981, p125) Mundanely put, voluntarism, in effect, stipulates that all coherent ampliative scientific belief-webs are rational. Yet, we may contrast the creative context of construction with the *context of* experiment or '*destruction*'. This is the context where our speculations are put to the test: they either succumb under the worldly resistance or they survive, at least for a while.

In the context of destruction, the evidential import is negative, whereas the relation of support between theory and evidence is up to the believer. For some, this sounds like surrender to inductive scepticism; others appreciate that we have to learn to live with our non-ideal epistemic situation. Objective warrant is not to be had, so we had better liberate ourselves from the illusion. Rich, elegant or unifying ontologies are constructed, but we have no reason to suspect they are, somehow, truth-conducive. Rich, elegant or unifying empirical systems are constructed likewise and we have no reason to suspect they are truth-conducive. The difference is that the empirical systems are vulnerable to the worldly evidence, while the non-empirical content, albeit truth-valued, is only vulnerable to the internal and ever shifting resistance of the scientific community.

## 2.6 Conclusion

The questions evoked by Putnam's intuition have been answered. The scientific success to be explained is empirical success with an emphasis on a wide, varied scope and the successful prediction of novel types of phenomena. This success is allegedly best accounted for by the truth of particular components of specific theories, which renders the vague notions of 'maturity', 'typical referential success' and 'approximate truth' superfluous. These explanations are intended to be the loveliest explanations that current science has to offer and we are plausibly supposed to infer their likeliness.

The envisioned conclusion (realism of some sort), however, is no neutral scientific hypothesis as it is *empirically* indistinguishable from the abduction to (likely) empirical adequacy. For example, if the empiricist believes that successful accepted science is empirically adequate, then he could also embrace the realist explanation—albeit only as being empirically adequate (accepting that the theoretical components responsible for the success apparently yield adequate empirical generalizations, will go on being successful and will be retained on a pragmatic basis). Finally, to believe an explanation may, from an epistemic view point, be similar to believing information *sec*.

This does not rid us of the deeply entrenched intuition that theories that enjoy empirical success over a wide and varied scope of phenomena and make

successful, empirical predictions of novel types of phenomena must, somehow, have something right about the unobservable. Many have the intuition that we look through the instruments and manipulate the microscopic objects. This is what science tells us, this is what we find in the textbooks and what we teach to the next generations. Our theories do not just formulate empirical generalizations, but the unobservable structures give rise to successful and sometimes surprising empirical generalizations. How can they not be approximately true? This is indeed the intuition, but iteration is not substantiation.

The inference to realism appeals to this intuition, but we still have no reasons to believe that the explanatory virtuousness of realism makes it likelier than not. As to why many people are realists, at least regarding the more mundane blood cells and microbes: these posits do arrange our ontologies in orderly fashion and help us to ‘summarize’ all empirical generalizations that spring from them. I am not sure whether the majority of working scientists is consciously a realist. The scientist might be a ‘working gnostic’, but it could be that she is not fully aware of the possibility of distinguishing epistemic and pragmatic commitment. *Even if* it turns out that most fully informed and philosophically-minded scientists embrace realism, this, of course, does not substantiate realism.

If the above diagnosis is correct, realists have no epistemic reasons to suppose that loveliness makes theories likelier to be true. That would mean that the conscious embrace rests, *ex hypothesi*, on a misunderstanding. Further reasons for such confusion are a subject for sociology or psychology. McAllister (1996, ch5), for instance, describes a sociological mechanism by which scientists tend to perform an aesthetic induction to the effect that some non-epistemic features of empirically successful theories are induced to be correlated with success. Loveliness might therefore be misjudged as having evidential import.

There is no reason to think that there is something epistemically special about manmade loveliness. A different ‘manual’ of internal criticism, e.g. the uglier the better, has, as far as we have reasons to believe, the same evidential status, although it may be rejected on pragmatic grounds. The same can be said of abductions to the *observable*, with the crucial difference that ampliative beliefs about the observable are not underdetermined by the permissible evidence: they have to face external resistance. “Since everything lies open to view, there is nothing to explain. For whatever may be hidden is of no interest to us.” (Wittgenstein [1953] 2009, 126)

If this account of explanation as purported information is combined with the responsible notion of scientific rationality as developed in the previous Chapter, and no other, non-empirical sources of scientific evidence are incorporated, then beliefs about the supra-empirical are rendered epistemically *reckless*: whereas the belief in empirical adequacy is vulnerable to experience, the beliefs beyond empirical adequacy are, at best, subject to ‘internal

resistance'. Since we have no reason to take this internal resistance as a genuine guide to truth, these beliefs are evidentially invulnerable, i.e. underdetermined by the permissible evidence, in which case there is no way in which their possible falsity can be exposed. Of course, non-empirical sources of evidence cannot be eliminated *in toto* on the basis of the above exercise, but loveliness seemed the most promising candidate and she can be put to bed.

## Chapter 3

# The Doxastic Policy

### 3.1 Introduction

Constructive Empiricism (CE) holds that the acceptance of a scientific theory involves as belief only that the theory is adequate to the actual observable entities, i.e. is empirically adequate (EmpAd). In the light of the aim of science, it is not needed to commit epistemically to claims concerning non-actuals or unobservables. CE's doxastic policy expresses this view of what is and what is not doxastically involved in theory acceptance. In theory, this is pretty straightforward, but in practice, the empiricist has to make judgments on actuality and observability to assess what his policy amounts to.

In the following, I tease out CE's proper policy on counterfactuals by considering a puzzle of actuality. Much has been written on CE's delineation of the observable, but the other criterion for being subject of empirical adequacy, *actuality*, is relatively neglected. I show that the limits of the actual are more restrictive than one might *prima facie* expect. A modally innocent account of counterfactual claims is available, but CE's doxastic policy renders the belief in counterfactuals scientifically supererogatory. It *seems* that CE needs to extend its policy so as to incorporate beliefs about the counterfactual to make sense of science and, notably, to delineate what can be observed. This is not the case: the original policy suffices, although the pragmatic dimension of theory acceptance, including the acceptance of associated counterfactuals, turns out to play an important role.

Two preliminaries: *first*, this Chapter is deliberately 'ocularist': I interpret un/observability as in/visibility, which allows me to say e.g. that invisible elephants are unobservable (even though they might be tangible and smell bad). The reason is that the literature offers a suitable semantic interpretation of seeing only. The ellipsis is obvious: observability in its proper sense is the disjunction of being visible and being accessible by the other sense modalities. *Second*, CE is a view of science and, while render-

ing epistemic commitment to non-actuals and unobservables supererogatory, it *does not prescribe* what we should or should not believe. Nonetheless, I sometimes abbreviate cumbersome locutions as ‘according to his doxastic policy, the empiricist need not believe  $p$  in the light of the aim of science’ by ‘the empiricist does not believe  $p$ ’ or ‘should be agnostic’ and the like.

## 3.2 A Puzzle of Actuality

### 3.2.1 Two Nuclear Power Plants

A nuclear power plant is needed. There are two competing blueprints on the table. According to our best scientific theories, the first, cheap one is likely to melt down; the expensive alternative is safe. To minimize potential harm, the second option is chosen and realized.<sup>1</sup>

How does CE deal with this scenario? We have a scientific theory  $T$  that covers meltdowns.  $T$  can be presented as a family of models with this kind of phenomena in its domain. A phenomenon is, by definition, an actual observable entity (object, event, process). We can describe a phenomenon by constructing a data model that represents it. A phenomenon fits inside  $T$  iff its data model can be embedded in one of  $T$ ’s models.  $T$  is EmpAd iff  $T$  has a model in which all the relevant phenomena (past, present and future) can be embedded. To accept  $T$  implies, according to CE, as belief only that  $T$  is EmpAd, i.e. that all the relevant phenomena fit inside  $T$ . Since the hypothetical meltdown of our example is avoided, this specific event is not actual and therefore acceptance of  $T$  does not imply the belief that the hypothetical meltdown fits inside  $T$ . How can the empiricist justify the choice for the safer power plant if this choice is based on beliefs that he considers, strictly speaking, supererogatory (not doxastically implied by theory acceptance, target for agnosticism)?

Our theory about meltdowns does not directly concern this *specific* hypothetical meltdown, but plausibly informs us about *general* circumstances and conditions that typically lead to meltdowns, i.e. it gives rise to claims of the form: *if* certain conditions are realized, *then* something will happen—and this is the general form of predictions. Thus, to accept  $T$  implies these

---

<sup>1</sup>To avoid problems with probabilities we could alter the example such that the meltdown is guaranteed. Thanks to F.A. Muller for suggesting this example to me. The scenario has interesting similarities with one discussed by Rosen (1994, p162 fn13), Ladyman (2000, p853; 2004a, p758), Monton and Van Fraassen (2003, p407–8) and Alspector-Kelly (2006, p374–7). Here is the gambit: suppose the empiricist is in the *unique* position to ‘bring into actuality’ a novel *type* of phenomenon . . . . The scenario is meant as an attack on CE’s portrayal of science at the motivational level: since science aims at empirical adequacy only, the empiricist allegedly has no scientific motive to *create* the phenomenon. Muller’s example with the power plants concerns a possible instantiation of a familiar type, but it raises the question whether the empiricist has a strictly scientific motive to *avoid* the phenomenon.

associated conditional beliefs and to believe their negation amounts to the rejection of  $T$ . The empiricist can uphold that his acceptance of  $T$  implies that he generally believes that certain circumstances lead to meltdowns, and additionally, he believes that building the cheap power plant would realize these circumstances. This is a prediction like any other.

However, fact is that in *this* case the circumstances are *not* actually realized, which renders the specific conditional ('if this cheap power plant were built, it would melt down') *counterfactual*—as becomes especially obvious in retrospect. So there is the rub again: although the statements that  $T$  gives rise to are general, the belief involved in accepting  $T$  pertains to the actual only.  $T$  is still EmpAd if it does not embed hypothetical happenings; the empiricist is (permitted to be) agnostic regarding  $T$ 's adequacy to these specific counterfactuals and he cannot retrospectively believe to have avoided potential harm within the confines of his doxastic policy.

### 3.2.2 Judging Actuality

This is a strange puzzle. If I were to ask the empiricist, "What happens if the cheap power plant is built tomorrow?" he should answer, "According to  $T$ , which I accept, it will melt down, and that is what I believe." On the other hand, if I ask him, "Do you believe that 'if the plant had been built, which it is actually not, it would have melted down'?" he passes over in agnostic silence.

The temporal dimension that emerges in this imaginary questionnaire does not concern actuality and therefore not EmpAd itself, but reflects a typical temporal asymmetry permeating our personal, fallible judgments of what is actual—in this case the actuality of the antecedents. There is an atemporal and mind-independent matter of fact as to what is actual and this determines the truth-values of the antecedents of the specific conditional claims that  $T$  gives rise to. If the empiricist had a crystal ball, he could restrict himself to beliefs with true antecedents and remain agnostic regarding any residue non-actualized possibilities.

We can think of  $T$ 's models as the possibilities that  $T$  allows. Acceptance of  $T$  does not require singling out the actual model beforehand; it just implies the belief that whatever (relevant) phenomena are unveiled, they all fit in one of  $T$ 's models. This is the courageous leap of faith in the face of the abyss of ignorance. However, to put his doxastic policy to practice, the empiricist should try to assess which claims are about actuality, and thus eligible for epistemic commitment, and which claims are not.

These personal judgments, from one's limited epistemic viewpoint, may be more appropriately phrased in terms of *defeasibility*. Consider an empiricist confronted with a specific claim of the form  $A \square \rightarrow C$ . The claim concerns observables only and follows from a theory that he accepts. Now, if he thinks that  $A$  is defeated, i.e. not true, he should be agnostic about the

claim, since his belief in the theory's EmpAd does not cover counterfactuals. If he thinks that *A* is not defeated, for example because it is a prediction, he may epistemically commit. The temporal asymmetry provides for an easy exposition, but defeasibility applies likewise to things past. He needs to judge the actuality of specific circumstances (past, present and future), to be able to believe what he can infer from his theory about these specific circumstances.

### 3.2.3 The Letter in the Drawer

Let us strip away the overt intrusion of intervention from the nuclear example. Then we are left with a bare counterfactual of this form: 'if you look in the drawer, you see the letter'—on the supposition that (i) the letter is inside the drawer, and (ii) you never actually look in the drawer.<sup>2</sup> Monton and Van Fraassen note about this example: "Even if the drawer is never opened the facts about its contents plus certain well selected generalizations about the circumstances suffice to determine anything about what would happen if the drawer were opened." (2003, p410) It is not clear how we are to reconcile this 'determining what *would* happen' with the claim that "empirical adequacy concerns *actual* phenomena: what does happen, and not, what would happen under different circumstances." (Van Fraassen 1980b, p60) Did Van Fraassen have a change of heart concerning the proper doxastic policy on non-actual events?<sup>3</sup> His comments are inchoate.

[M]any counterfactual conditionals are straightforwardly implied by strict logical implications, or by logical implications relative to theories or even simple statements the speakers take for granted. An example would be 'If you had looked in that drawer you would have seen the letter', which the speaker would justify with 'Because the letter is in that drawer', fully confident of various generalities about vision, light, and the like. (Monton and Van Fraassen 2003, p410)

Some accepted toy theory might give rise to the generalisation that 'for all letters and all drawers: if some letter is in some drawer and someone looks in the drawer, he sees the letter.' Combined with the first supposition—the letter is in the drawer—this yields ('relative to theory and simple statement') the conditional 'if you look in the drawer, you see the letter', but given the second supposition—you never look in the drawer—the antecedent is false and the empiricist should, again, remain agnostic.

<sup>2</sup>This example is borrowed from Monton and Van Fraassen (2003, p410). The second supposition is, although somewhat artificial, given and it is important for our example that the empiricist, *ex hypothesi*, believes this.

<sup>3</sup>At least initially, it was explicitly stated that, as far as science is concerned, we need not believe anything about non-actual entities. Textual evidence is abundant: Van Fraassen (1980b, p3, p13, p60, p64, p197, p202–3; 1989, p71, p91–3, p213–4, p228).

To phrase this objection slightly more fitting to the semantic explication of EmpAd, we follow Muller (2005), who gives a metaphysically innocuous interpretation of subjunctive conditionals (*were-would*) and explicates the ‘confidence of various generalities about vision, light, and the like’ in models of the wave theory of light, thereby fixing certain circumstances in a non-arbitrary and scientifically informed manner. Let  $T$  be the family of models that are allowed by the wave theory, presupposing room temperature and pressure, and representing the letter, your eyes, a light-source, the electric field of this light and a four-dimensional, spatiotemporal environment. (*Ibid.*, p74) The letter is given an appropriate size and reflection spectrum, the indexical ‘you’ is fixed as eyes with average sight, the light source is bright, the environment is your typical office. All is spelled out properly and elaborated upon by Muller.

Let  $\text{Front}(\text{letter})$  summarize that you look in the drawer, at rest, from an appropriate distance, and that the letter is inside the drawer.  $\text{See}(\text{letter})$  is the claim that you see the letter. (Sentences are related to the models by an interpretation-function and truth-conditioned by an interpretation in the classical sense.) Now, the subjunctive conditional ‘if you had looked in the drawer you would have seen the letter’ means relative to  $T$ : for all of  $T$ ’s models wherein  $\text{Front}(\text{letter})$  is true,  $\text{See}(\text{letter})$  should also be true, and  $T$  has at least one model in which it is true that  $\text{Front}(\text{letter})$ . (*Ibid.*, p96) In symbols:

$$\forall \mathcal{M} \in T : \text{tr}(\mathcal{M}, \text{Front}(\text{letter}) \rightarrow \text{See}(\text{letter})) \wedge \\ \exists \mathcal{M}' \in T : \text{tr}(\mathcal{M}', \text{Front}(\text{letter}))$$

All models of  $T$  where either the letter is not in the drawer or you do not properly look in the drawer make the material implication vacuously true; the important thing is what happens when you look at the letter. We can construct at least one such model whereof  $\text{Front}(\text{letter})$  is true, satisfying the second operand of the conjunction. Furthermore, it seems that for all models in which you look at the letter (with abovementioned qualifications) you also see it, satisfying the first operand. Hence, the subjunctive conditional is true relative to  $T$ .

So, we have a modally innocent interpretation and the claim is true relative to the wave theory of light. Suppressing above provisos, the conditional amounts to the following claim: ‘relative to the wave theory of light: if you had looked in the drawer you would have seen the letter.’ If you accept the wave theory and have no beliefs relevant to the antecedent, you could believe the non-relativized ‘if you had looked in the drawer you would have seen the letter.’ Regrettably, by supposition (ii), we know that none of the models that satisfy the antecedent embeds an actual event. As Muller reminds us, subjunctive claims “fall unproblematically under the purview of the epistemic policy of CE: believe only those accepted modal propositions

(i.e. modal propositions relying on accepted scientific theory) that are about actual observables only and remain neutral *qua* belief about all other accepted modal propositions.” (2005, p97)

Things get even worse when we start focussing on explicitly tensed or very fine-grained claims and the identity and individuation of entities over time. The models of a theory are vulnerable to whatever may empirically present itself; the theory lays out a multitude of possibilities and then we hope that the phenomena are ‘caught’ in the web. The epistemic commitment to EmpAd, by contrast, does not encompass all these non-actual possibilities. Suppose, you actually look in the drawer and, indeed, you see the letter. Now, what would have happened if you had looked a minute ago—which you actually did not? The empiricist shrugs; the event of you looking in the drawer a minute ago is counterfactual, thus beliefs about it are supererogatory. And what if you had looked a second ago? Or if you had looked at the exact same time, but meanwhile breathing out instead of breathing in? Well, this depends on how ‘fragile’ or ‘robust’ events are construed. (Cf. Lewis 1986b, p255–9) This is not the place to delve into the essences and mereology of events. Yet, it would not be good if the empiricist cannot deny—as far as science is concerned—claims of the form, ‘if I had blinked my eye, you would have evaporated,’ whenever he believes that you did not blink.

### 3.3 On What is Actual

#### 3.3.1 Against Type-Actuality

We could try and tinker with the limits of the realm of the *actual*. Then, *prima facie*, two escape routes appear—neither is viable. The first route, to type-actuality, is dealt with here, the second, the resort to object-actuality, is examined in the next Subsection.

The first option is to promote the property of being actual to a type-level. The other criterion for belief, i.e. being about the observable, can be thought of as concerned with a type-level property: apples (in general) are observable and if some  $x$  it not observable, then it is not an apple. For actuality this does not fly. It is not the case that apples are generally actual: specific apples are actual, others merely hypothetical. This difference, however, is obscured by the often mentioned paradigm case of a non-actual entity: the flying horse. “[T]he distinction [...] between *racehorse* and *flying horse* [is clear]: the first corresponds to something in the actual world, and the other does not.” (1980b, p82)<sup>4</sup> Confusingly, ‘flying horse’ is a *type* of non-actual entity and ‘racehorse’ is a *type* of actual entity, but not every possible racehorse

---

<sup>4</sup>Also see Van Fraassen (1980b, p15), Monton and Van Fraassen (2003, p420) and Muller and Van Fraassen (2008, p202). The other example of a non-actual entity is the ride of the headless horseman (Van Fraassen 1980, p197).

corresponds to something in the actual world: I am bereft of a racehorse, so ‘my racehorse’ refers to an imaginary friend (and my claims about it are counterfactual).

The sting of our puzzle is quietly sidestepped if the actual is treated on a type-level, and sometimes terminology is clouded by switching to the ‘real’ versus the ‘fictional’ or to ‘the actual world’ in general.<sup>5</sup> Before blocking this route, let us briefly see how it could help CE. Let a type of entity be *t-actual* iff it has at least one instantiation *a* in the world. This t-actuality is inherited by all possible tokens of the t-actual type. Racehorses are t-actual, flying horses are not. If *a* is observable, we have, in principle, epistemic access to it and we can test our theory by comparing the theoretical model to *a*’s data model. If our theory is adequate to *a*, we may believe it is adequate to every possible token of the associated type. Acceptance of *T* implies as belief that all relevant t-actual observable entities are saved by *T*. Since at least one meltdown is actual, *T* is believed to be adequate to all possible meltdowns, and therefore we can believe that ‘if the plant was built, the meltdown would have occurred.’

Alas, aside from leading to cumbersome talk of non-actual t-actual entities, this construal of acceptance violates the empiricist spirit for at least three, related reasons.

(A) Some (most?) empiricists are modal anti-realists: minimally, they reject that the world offers objective truth-conditions for statements about possibility and necessity. As they see it, our claims about non-actual entities may be true about some model, but the non-actual content of that model is not (directly) related to the world. To believe *p*, as opposed to merely accept *p*, is purportedly to believe that *p* is true about the world. So our considered extension of CE’s doxastic policy will prescribe beliefs, namely those about t-actual non-actual entities, which the modal anti-realist cannot believe.

(B) The adoption of t-actuality demands a substantial typology of ‘the possible’ to ground the doxastic policy. The nominalist can maintain that, although we cut up the world in *some* way, we need not claim that we cut it at its natural joints. Theories can classify phenomena in one way or another, but not much depends on this, as long as the phenomena are saved. Whether they are scientifically in need of saving is unrelated to our taxonomic choices: we can just wait (or act) and see what happens, and whatever happens needs to be saved.<sup>6</sup> This changes with the introduction of t-actuality. Whether the claims about my imaginary racehorse need to be saved by science, whether they are covered by EmpAd, and whether the empiricist should believe them—all depends on how this non-actual entity

<sup>5</sup>Van Fraassen (1980, p82; 1985, p256), Muller (2004, p646).

<sup>6</sup>CE’s policy does rely, in practice, on one taxonomic choice: the delineation of the observable. Although the domain of the observable is not itself theory-dependent, the empiricist should have a theory-informed judgment of the limits of the observable, if he wants to put his policy in practice.

is classified: as an object, an *Equus ferus*, or as a token of the type my racehorse? These choices cannot be made at a whim: our interpretation of the aim of science depends on them! Hence, the retreat to t-actuality commits the empiricist to a thick ontological classification.

(C) CE's doxastic policy is, at least partly, motivated by an empiricist conception of the limits of our epistemic reach. Epistemic commitment (belief) is limited to claims concerning the actual observable. These beliefs are empirically vulnerable: they are able to conflict with future experience. If we want to apply our models to non-actual entities, this conflict is simply no option. It may be retorted that claims about my t-actual horse are *indirectly* put to the test via the actual instantiations of the relevant type, but the fact remains that the models are invulnerable to the particularities of my fictional friend: I imagine my horse to be a white Lipizzan, but my neighbour strongly disagrees and has a Przewalski's horse in mind. How to decide? These disputes cannot be settled scientifically. We conclude that what is actual depends exhaustively on the wordly particularities and should not be promoted to a type-level.

### 3.3.2 Against Object-Actuality

The second route is even less promising. CE could redefine 'the actual' to encompass all possible events and processes involving actual *objects*. Actuality is then, so to speak, grounded in objects and bubbles up to any possible interactions featuring these objects. By redefining 'the actual', the empiricist could claim of the letter in the drawer that, since the associated counterfactual is wholly about actual objects, and additionally these are observable, it is relevant to EmpAd and subject for empiricist belief.

There is something appealing to this escape route, for the letter-drawer-counterfactual seems to depend on how the world actually is, including the actual letter in this actual drawer. Also, this route rids us of most of the problems with explicitly tensed claims about events. When trying to put the doxastic policy in practice, I can ignore whether some conditional claim is about a second ago, in which case the antecedent is settled, or about a second from now, and not yet defeated. Moreover, there is something very counterintuitive about a situation in which we can have all kinds of elaborate and specific beliefs about the drawer and the letter, but must remain agnostic only about the consequence of the non-actual event of you looking in the drawer. If we make actuality dependent on objects only, we can infer from these specific beliefs what would happen under different circumstances and believe that as well.

Alas, this route is no less problematic. Note that the retreat to object-actuality does not solve our initial nuclear puzzle: the cheap reactor is never built, so we cannot piggyback the meltdown on the reactor's actuality. Moreover, this construal inherits two of the problems of type-actuality: it forms

an incoherent mixture with modal anti-realism and is orthogonal to the empiricist spirit: “Make your requirements of verifiability or testability as weak, as modest as you like, and they will still never allow you to say: we have checked that this model is correct about what did not happen, about the experiment we did not carry out, as well as about what did happen.” (Van Fraassen 1989, p92–3)

## 3.4 Easing the Doxastic Policy

### 3.4.1 The Revised Policy

CE’s doxastic policy has already undergone one revision: a concession allowing some minimal beliefs about the non-actual. Unfortunately, this revision is of no help to our puzzle. I briefly rehearse the dialectics that prompted the revision. Musgrave (1982, p266; 1985, p208), and Maxwell before him (1962, p9–10), noticed that if you do not believe claims about the unobservable, you cannot very well believe that some  $x$  is unobservable, even if your best scientific theory  $O$  says so. Now you are in trouble if you try to explicate what your doxastic policy amounts to. On a side note: the same applies to non-actuals: ‘ $x$  is non-actual’ is not a claim about an actual entity.<sup>7</sup>

Van Fraassen (1985, p256) initially replied that acceptance of  $O$  implies the belief that all the relevant phenomena are saved by  $O$ , so if some relevant entity is not saved by  $O$ , we believe that it is apparently unobservable if actual. Equivalently, I add, the relevant entity is non-actual if observable. Since the EmpAd of  $O$  covers all and only actual observable entities within  $O$ ’s domain, we may infer and believe that there are no relevant actual observable entities overlooked, i.e. any relevant  $x$  not embedded by  $O$  is not-(actual  $\wedge$  observable). We cannot attribute categorical unobservability, nor categorical non-actuality. The doxastically relevant entities are, so to speak, delineated from inside out.

Driven by the intuition that we should be able to predicate unqualified unobservability, Muller and Van Fraassen (2008) proposed to extend our beliefs about un/observability to the realm of the possible.<sup>8</sup> “If you accept T, and  $Y$  is (un)observable according to T, then believe so.” (2008, p204) After this conceptually prior phase, we return to the original policy, but leave the newly gotten un/observable distinction untouched. Now we can distinguish three of the four types of entity: (i) racehorses were considered unproblematic, (ii) flying horses are observable, and since accepted biology does not model them (we have never seen one), they are non-actual, (iii) invisible

<sup>7</sup>Neither is “The ride of the headless horseman is an observable event, but not an actual one.” (Van Fraassen 1980b, p197). Van Fraassen may be talking *ex cathedra*, but he should not believe this. Cf. Muller (2004, p646).

<sup>8</sup>For the discussion leading up to the revision, see Muller (2004; 2005) and Dicken and Lipton (2006). For comments, see Dicken (2009a; 2009b; 2010, 3.2).

flying elephants are unobservable and therefore we suspend judgment as to whether they are actual or not.

Notice that the non-actuality of ‘flying horses’ is concluded on a type-level (ii). Our theory of observability leads us to conclude that meltdowns are observable, but since our nuclear theory *does* model meltdowns, this should imply nothing about specific counterfactual meltdowns, on the penalty of promoting actuality to a type-level. Still, if biology does not allow for models representing the type flying horse and Pegasus turns up, biology is not EmpAd. Thus, accepting biology implies the belief that all possible flying horses are non-actual, while it does not imply the belief that all possible race-horses are actual. There is an asymmetry between the doxastic implications of the presence and the absence of a model-type of an accepted theory.

The revised policy thus licenses beliefs concerning the non-actual, e.g. that flying horses or hypothetical meltdowns are observable, but only insofar this belief is restricted to a judgment of un/observability: thereafter the original policy is put back to play. Since this original policy recommends agnosticism concerning claims that transcend actuality, the revision does not solve the nuclear puzzle: although the empiricist now classifies the non-actual meltdown as an observable entity, the meltdown still does not satisfy the conjunction (actual  $\wedge$  observable), it is not covered by EmpAd and the belief that the meltdown would have occurred is, thus, supererogatory. It seems that to solve the puzzle, we need to have additional beliefs about non-actual entities. In the case of the cheap power plant, notably, that the meltdown would have occurred. Before considering the option of relaxing CE’s doxastic policy even further, I turn to one final problem, underlying this revised policy.

### 3.4.2 Counterfactual Observations

Observability is relativized to our epistemic community; something is observable iff it is observable-for-us humans, who can be understood as measuring devices from the physical point of view. Once our community is specified, the limits of what is observable are fully dependent on the world. (Van Fraassen 1980b, p16–8, p57; 1985, p305) However, to put her policy to practice, the empiricist has to judge what is observable. The limits of observability are a theory-independent matter of fact, but our assessments of these limits cannot proceed from the armchair; they are guided by science.

Van Fraassen offers a rough guide to judge the limits of what we can observe:

X is observable if there are circumstances which are such that, if X is present to us under those circumstances, then we observe it. (1980b, p16)

The ‘we’ signifies a ‘normal’ representative of our community.<sup>9</sup> Let us interpret the right hand side of the guide as not only implied by but also implying observability, so that it doubles as a guide to conclude unobservability by contraposition. The implication on the right hand had better be read as a subjunctive implication; otherwise  $X$  is vacuously observable whenever it is not present under some circumstances. This yields the following rough guide, which is implicitly relativized to our epistemic representative (such that Present reads present-to-us and Observe observed-by-us) with  $c$  standing for—as of yet mysteriously unqualified—circumstances.

$$\text{Observable}(X) \equiv \exists c (\text{Present}(c, X) \square \rightarrow \text{Observe}(X))$$

The drawback of this explication is that it signals a commitment to counterfactuals similar to the letter-drawer-example: there are circumstances such that if  $X$  were present, we *would* observe it.<sup>10</sup> Additionally, we need a principled way of selecting permissible circumstances. If our judgments of an entity’s un/observability depend on the possibility of observing it, this, presumably, concerns some notion of physical possibility: it is logically possible to observe almost anything.

The reproach to Van Fraassen, initially by Rosen (1994) and taken up by Ladyman (2000), is that he needs to believe in claims that transcend actuality to explicate CE’s original policy, while this same policy recommends agnosticism about claims that transcend actuality. This tension is inherited by the *revised* policy. The revised policy relies on a conceptually prior delineation of the observable, such that this delineation is itself not subject to the policy. (The delineation had better be excluded from the policy, for it cuts across *all* observables, actual and otherwise.) However, if assessments of observability are based on the rough guide, the empiricist seems committed to believe in the truth of claims about counterfactual observations. This does not accord well with the empiricist spirit.

To illustrate the tension, we borrow Muller’s (2005, p83) example of the claim that should render dinosaurs observable. As intimated, Muller proposes to relativize modal claims to the models of a relevant theory, in this case the wave theory of light, thereby principally fixing the permitted circumstances. Additionally, he advances an entirely extensional rendition of the rough guide.

---

<sup>9</sup>Cf. Van Fraassen (1980b, p17). As Muller remarks, normality is to be explicated by various experts (2005, p63). Our epistemic community  $\mathcal{E}$  might evolve or we may decide to extend it to cover other beings (Van Fraassen 1980b, p18), so in the future we might speak of observability and consequently phenomena and EmpAd relative to  $\mathcal{E}_{2013}$  or  $\mathcal{E}_{3001}$ . Preferably, a theory saves the union of all thusly relativized phenomena.

<sup>10</sup>Discussion of this problem can be found in Rosen (1994, p169-76), Psillos (1999, p189–90), Ladyman (2000; 2004a; 2011, p95), Monton and Van Fraassen (2003), Hanna (2004), Muller (2005), Ladyman et al. (2007, ch2) and Dicken (2007; 2009b; 2010, 3.3).

$$\text{Obs}(X, \mathcal{E}, \mathbf{L}) \equiv \forall p \in \mathcal{E}, \exists \mathcal{M} \in \mathbf{L}: \text{tr}(\mathcal{M}, \text{Front}(p, X) \wedge \text{Sees}(p, X))$$

In natural language, this reads: object  $X$  is observable for normalized community  $\mathcal{E}$  relative to the set of models  $\mathbf{L}$  iff for every member of the community there is a model in  $\mathbf{L}$  in which it is both true that the person is in front of the object and he sees it. Applied to a properly modelled 3D Littlefoot<sup>11</sup>, relative to our community and a set of models structurally similar to those of the letter-drawer-example, we establish that there is, indeed, a model, e.g. representing the observer standing right in front of the dinosaur in broad daylight, in which every member sees the dinosaur.

This definition is confined to objects and Littlefoot is allegedly an actual observable *object*. The problem is, however, that plausibly none of the models satisfying the definition represents an actual *event*. As long as the empiricist does not resort to object-actuality, none of the models is an ‘actual model’: it represents a non-actual entity, it cannot impugn on EmpAd, and it is invulnerable to experience.

The empiricist has good reasons to suspect those models of representing non-actual events. Regarding events involving some golden sphere with a substantial diameter, the empiricist might claim that the actuality of ‘the antecedent’—in this case redirected to the model—is not yet defeated and his beliefs about it are like any empirical prediction. (Whether his assessment is correct depends entirely on the actual world, but he should practice his policy despite his ignorance about the limits of actuality.) In case of the dinosaur, by contrast, the empiricist plausibly judges ‘the antecedent’—the representation of someone being in front of the dinosaur—as being counterfactual. According to the *original* doxastic policy, he should therefore not believe the claim that Littlefoot is observable.

Ladyman (2004a, p760) wonders why and whether the empiricist is justified to counterfactually hold fixed certain generalizations. Why not suppose that the dinosaur would turn invisible every time one hypothetically gets in its proximity?<sup>12</sup> Can the empiricist insist that light will maintain its familiar behaviour in non-actual circumstances? The theory of light is firmly

---

<sup>11</sup>



<sup>12</sup>The empiricist could retort: well thusly modelled, the dinosaur is unobservable, but

grounded in beliefs about the actual world, but if the empiricist wants to claim that counterfactual luminous misbehaviour is *impossible*, it seems he is committed to believe in the truth of this modal claim. These questions also pertain to the extensional definition: they can be interpreted as asking for a justification of the modelling choices.

The resulting dilemma is this: with the *revised* policy in place, the empiricist divides entities into observables and unobservables regardless of their actuality. But if this dichotomy is based on models of non-actual observations, the policy gives the impression of a halfway house. The models of the theory of light provide a partial context of physical possibilities, particularly the possibilities of observing. The empiricist applies this context to non-actual circumstances. This seems to commit him to the truth of an idiosyncratic subset of counterfactuals, i.e. those covered by the general claim that the models of  $\mathbf{L}$  are *counterfactually adequate*. However, if the empiricist retreats to his *original* policy, the limits of the observable, if spelled out in terms of possible observations, can only be believed in terms of actual observations. Then, CE reduces to *manifestism*: the view that science aims at manifest adequacy, where ‘the manifest’ is comprised of all entities that we actually observe—past, present and future.

### 3.4.3 Incorporating Non-Actual Observables

Most of the above problems are solved in one fell swoop by further relaxing the doxastic policy so as to recommend beliefs about all possible observables, where the relevant notion of possibility is constrained by the models of physics. The definition of observability is exempted from the policy: observables are defined in terms of possible observation-events, so if these observation-events need to be observable themselves to be allowed by the policy, we are caught in a regress. Therefore, the definition is not subordinated to the policy—since the policy contains the term ‘observable’, this simply comes down to the stipulation that the policy is not self-applicable. The policy and the definition of observability are more harmonious than in the original situation.<sup>13</sup> With the *extra extended policy*, we can conclude

---

I reject that model. I believe that the adequate model of the dinosaur has such and such properties, which render the dinosaur observable and *that* is, admittedly, a leap of faith. This response only partially succeeds, for the object may be actual and therefore ‘covered by’ EmpAd, but the event of ‘getting close’ is non-actual and its consequences need not be embedded.

<sup>13</sup>This off-hand exemption may seem more problematic than presented. The policy governs the scope of the epistemic commitment that is scientifically ‘needed’. It implicitly relies on the definition of observability which is, in turn, unpacked as an epistemic commitment, i.e. to claims about possible observations, which cannot be ‘validated’ by the policy on pains of an infinite regress. By failure of self-governance, the policy renders itself scientifically supererogatory. However, such a feature is not always disastrous: Popper’s scientific criterion of falsifiability cannot be applied to itself and is therefore, by its own lights, not a scientific requirement—it is rather a philosophical principle. We can similarly

categorical unobservability and remain agnostic about the actuality of unobservables. Imaginary horses, counterfactual meltdowns and the event of looking at the letter in the drawer are all observable and, consequently, the subject of empiricist beliefs.

Before rejecting this modification, let us see what it amounts to. Keep in mind that CE has a teleological and a doxastic component. The latter states, in effect, that we need to believe theories only insofar as they contribute to the aim of science, which was EmpAd. This yields an elegant, coherent and encompassing view of science. The extra extended policy recommends beliefs about all possible observables, so it stands to reason that we extend the aim of science accordingly. To illustrate this, let us consider the schizophrenic case of the modal realist—asserting that there are theory-independent facts about the non-actual—who holds, nonetheless, that science aims at EmpAd only. *Surprisingly*, Monton and Van Fraassen suggest that the combination of modal realism and original CE is quite unproblematic.

[O]ne can believe what our best theories say about observable entities (whether actual or non-actual), but not what our best theories say about unobservable entities (whether actual or non-actual) [...] It would thus be natural for a person who is a constructive empiricist and modal realist to believe what our best theories tell us about flying horses, but to be agnostic regarding what they tell us about the number seventeen. (2003, p420, my emphasis)

However, this ‘naturalness’ is misleading. The combination states that there are objective modal facts and truth-conditions in the world, but it is not part of the aim of science to ‘save’ them; the aim of science can be fulfilled while remaining silent about modal facts, and modal beliefs (perhaps in contrast to the mere acceptance of modal statements) are scientifically supererogatory.

Admittedly, the cocktail is logically consistent, but hardly recommendable. This hybrid philosopher has to judge her own beliefs about flying horses and counterfactual meltdowns as unneeded in the light of science’s aim. Regarding the power plant example, she should say that she believes that the cheap option would have been disastrous (modal realism), although this belief is scientifically superfluous (CE). From her point of view, this would make science an arbitrarily limited endeavour. She believes to have epistemic access to the world, whilst parts of the world may be epistemically ignored by our paradigmatic epistemic investigation.

To solve the tension, and to make counterfactuals on meltdowns scientifically *non-supererogatory*, the aim of science should accord with the beliefs involved in theory acceptance. Both components of CE should be modified, such that science aims at adequacy to the observable—actual and

---

say that the extra extended policy is not intended to be intra-scientifically ‘needed’; it is a meta-scientific characterization of what is doxastically involved in science.

otherwise—and acceptance of a theory implies as belief only that it serves this aim.

One need not necessarily convert to modal realism to support this policy. We could just say that claims about possible observables are made true by our accepted models. But why—and more importantly: to which extent—should we put faith in our models? I might construct some model of my imaginary Middle-earth, but I will not persuade many to *epistemically* commit. What would such a commitment purport? Usually, believing  $p$  means believing that  $p$  is true about the world. If beliefs about non-actual observables are not supposed to be true about the world, then I am not sure what such a claim amounts to—other than an *as-if*-commitment comparable to the assertion that ‘Sherlock Holmes has a cap’ is true in the book. If the models are supposed to derive their status from their testability against *actual* observables, then we are back in the conundrums of type-actuality. What is the status of precisely those models that do *not* embed actual entities, if these are not truth-conditioned by the world? If modal realism is denied, we may wonder *what it means* to commit epistemically to model realism.

Irrespective of the semantic elaboration (modal or model), opening the flood gates to all possible observables is, as with the flight to type- and object-actuality, detrimental to the empiricist spirit. The choice for EmpAd, instead of e.g. truth, is motivated by the demand for direct empirical vulnerability. Claims that do not meet this demand may be deemed vulnerable, but this is only and always by means of the empirical import that can be inferred from them. They are unable to meet the tribunal of experience autonomously. In other words: the supposed indirect vulnerability equals, if taken in isolation, experiential invulnerability. This applies to all supra-empirical claims: to claims about unobservables and equally to claims about counterfactual observables.

## 3.5 A Humean Dissolution

### 3.5.1 Committing to Counterfactuals

According to CE, theory acceptance has a pragmatic face. I have downplayed this dimension for ease of exposition, but it is time to bring its importance to the fore. Acceptance implies as belief only that the theory is EmpAd, but alongside this strictly epistemic dimension, it involves a commitment to immerse oneself in the theory and act, perhaps even verbally act, as if the theory was true.

“The depth of commitment is reflected [...] in how the person is ready to answer questions *ex cathedra*, using counterfactual conditionals and other modal locutions, and to assume the office of explainer” where

“scientific explanation is not (pure) science but an application of science.” (Van Fraassen 1980b, p202, p156)

This pragmatic commitment allows the immersed empiricist to say, even in retrospect and *contrary to fact*, that if the cheap power plant had been built, the meltdown would have occurred. This claim is not part of pure science, but an application of it; a pragmatic extension to hypothetical scenarios. Pretence is not belief. If we seriously inquire into the scope of his doxastic attitudes, for instance in the context of a philosophy class room, the empiricist should admit agnosticism regarding the counterfactual meltdown. He can only uphold a sincere epistemic commitment regarding conditional claims as long as he judges the antecedent as actual. In the case of the meltdown, he should retrospectively ‘admit defeat’, and switch to agnosticism.

This Section opened with the question how the empiricist could scientifically justify the choice for the safe plant in retrospect. *Sotto voce*, he cannot. Models of counterfactual meltdowns are not vulnerable to experience and neutrality is the proper attitude. . . in the philosophy class room. In there, he might say that his pragmatic commitment to the counterfactual is *based on* his epistemic commitment to nuclear theory. As a practicing and immersed scientist, by contrast, he simply avows the claim. If he is committed, it might even be pragmatically incorrect not to.

Regarding the letter in the drawer, I might add that part of the counterintuitiveness of the scenario springs from the stipulation that you never actually look in the drawer. This is an artificial example. In real life, one will seldom judge this antecedent as tenselessly defeated. Without the stipulation, the empiricist *should* straightforwardly believe the claim whenever the assumed EmpAd of the relevant theory implies it. When explicitly tensed—‘if you had looked in the drawer a minute ago, you had seen the letter’—the nature of his commitment depends on his assessment of the antecedent. If the empiricist believes its negation, he may, again, withhold epistemic commitment to the conditional claim. His epistemic commitment is restricted to claims about actuality and need not extend to ‘what if’-stories, which are not truth-conditioned by the world. Nonetheless, he had better reserve this reflection for the class room. In everyday life it would be quite insensitive and pragmatically ridiculous to admit agnosticism about the tensed letter-drawer conditional.

This emphasis on the pragmatic commitment of the immersed scientist should, however, not distract from the empiricist view of science. In the light of the aim of science, it is not necessary to commit epistemically to claims about events or processes that did not happen or objects that do not exist.

### 3.5.2 Judging Actual Observability

A *pragmatic* commitment to counterfactuals, notably those concerned with observations, guides our judgments of the observable. There is, however,

some unease with a situation in which the practical import of the doxastic policy is based on non-epistemic commitment, and our judgments of the observable are no part of ‘pure’ science, but a ‘mere’ application of it, while they determine what counts as pure or applied. Therefore, the guidance offered by the pragmatic commitment to counterfactual observations, should, at bottom, be spelled out in terms of an epistemic commitment to claims about actuality.

The situation is not as disastrous as one might think; recall that the un/observability of some entity depends exhaustively on the features of this entity, once our epistemic community is specified. The theory-independence of the un/observable prevents the dichotomy from sliding into a non-epistemic *as if*-distinction. Still, the dichotomy, and, consequently, the aim of science and the doxastic policy are *practically* empty if we have no way of judging observability—and severely crippled if our judgments of the observable are reduced to beliefs about the manifest.

The moons of Jupiter are not yet observed with the unaided eye. Our detection of them with the help of instruments may be counted as observation in virtue of the supposition that if we *were* close enough, we *would* see them (with the naked eye). When I confront the empiricist in the class room with the associated observation-conditional, his attitude depends on his assessment of the antecedent. If he judges the antecedent to be false (we do not actually get close enough), and hence, the conditional as counterfactual, he may, *sotto voce*, admit agnosticism. On the other hand, he might commit pragmatically to the (to his mind) counterfactual observation-conditional, based on his acceptance of the theory of light and a suitable model of the moons. Now, if he, subsequently, assesses the limits of the observable, to spell out the consequences of his doxastic policy, should he then rely only on his bare epistemic commitment, excluding the moons, or additionally on his immersed *as if*-commitment, plausibly including them?

Dicken claims that a ‘committed agnosticism’ to counterfactual observations, comparable to the acceptance of claims about unobservables, suffices to determine the dichotomy. (2007, p609) This ‘committed agnosticism’ is spelled out as the combination of literal semantics, doxastic agnosticism and some additional kind of commitment. However, he misrepresents the nature of this commitment: “when the constructive empiricist accepts a theory, the attitude he takes towards statements about unobservables involves [...] a substantial *epistemological* commitment.” (*Ibid.*, p608, my emphasis) This is not correct. The whole point of CE is that one may refrain from epistemically committing to supra-empirical claims. As with claims about unobservables, acceptance of counterfactuals does not involve a ‘substantial epistemological commitment’.<sup>14</sup>

<sup>14</sup>Dicken considers both the agnosticism and the commitment as epistemic, which, to my mind, renders ‘agnostic commitment’ an oxymoron and leads to such puzzling claims as “whether or not the constructive empiricist believes that [the] truth-conditions are met

The pragmatic commitment to counterfactuals does point towards a solution. Why suppose that we would see Jupiter's moons if we were close enough? The intuition behind the conditional concerns the judged *actual* features of these moons. Monton and Van Fraassen claim that the property of being observable is comparable to the property of 'made of brick' or '75 feet long'. (2003, p413) Alongside his metaphysically innocent construal of the subjunctive conditional and the extensional definition of observability, Muller (2005, p81) proposes a scientific criterion. The criterion, for the visibility of objects, is in the spirit of ' $\geq 1\text{cm}$  long': it concerns size and reflection spectrum only.

We should not think of these crucial features as making true claims about counterfactual observations. We turn it around, and interpret the counterfactuals as *thought experiments* that may guide us to predicate the relevant features to unobserved entities, including the non-manifest, thereby assessing the limits of the observable.<sup>15</sup> To accept a model of the moons of Jupiter or to accept a model that renders the non-manifest dinosaur observable requires, indeed, a leap of faith. This leap is just the leap towards EmpAd. The courage typically involved in theory acceptance is also needed to assess the limits of the observable. The practical delineation does not essentially depend on the pragmatic commitment to counterfactuals, but this commitment may be an instrument to judgments of the crucial features involved. These features, and observability itself, are not counterfactual and judgments of them are judgments of the actual.

The scientific criterion should be extended to cover the other entities and the other senses in similar vein and is, contrary to the theory-independent realm of the observable itself, not immune to revision. It helps to put the doxastic policy to practice, and it can even be subordinated to this policy, since it does not depend on counterfactual observation-events. The *revised* policy enables us to classify any possible entity as un/observable, but methodological continuity might promote dropping this revision and revert to the *original* policy, to the effect that the actual observable is delineated from inside out and establishing unqualified unobservability or the alleged observability of Pegasus and other non-actual entities is not deemed part of the aim of science, although it may be subject to pragmatic commitment.

### 3.5.3 Janus-Faced Commitment

We reject the option of extending the epistemic commitment to allow for beliefs about non-actual observables, but only at the price, it seems, of

---

is irrelevant, provided he is suitably committed to the claim that they are." (*Ibid.*, p609)

<sup>15</sup>Since the scientific criterion is co-extensive with the extensional definition (*Ibid.*, p84), we might as well use the definition for our judgments of observability, but we should keep in mind that we need not believe the counterfactual applicability of the models: it is only the final verdict that counts.

severely blurring the distinction between the epistemic and the pragmatic commitment. If the immersed empiricist should act as if he believed the supra-empirical claims and even avows them in all contexts but the classroom, both types of commitment coincide, at least outwardly, for almost all intents and purposes. This may lead one to think that there is no real distinction to be made.<sup>16</sup>

The charge that a pragmatic commitment cannot be distinguished from belief surprises me, for I think the distinction is quite straightforward and commonsensical: we are all familiar with the joys of make-believe. However, the terminology encountered is a bit of a hodgepodge: we believe, accept and commit, based on several opposing virtues: confirmational versus informational and epistemic versus pragmatic.<sup>17</sup> Most confusion will be solved by regimenting the vocabulary. The acceptance of a theory implies both an epistemic and a pragmatic commitment. The opposition of belief *versus* acceptance, when applied to a theory, is confounded: belief is part of acceptance.<sup>18</sup> Belief is the epistemic commitment; the pragmatic commitment could be called *mere* acceptance, contrasted to *full* belief. The opposing pairs of virtues are illuminated shortly.

Van Fraassen repeatedly stresses that acceptance does not collapse into belief across the board, since there are non-epistemic reasons for theory acceptance: “If some reasons for acceptance are not reasons for belief, then acceptance is not belief. And indeed some reasons for acceptance hinge crucially on the audacity and informativeness of the theory. So acceptance is not belief.” (1985, p281) This is both *confused* and *confusing*. It is confused, since the fact that we can have different reasons for adopting some attitude does not establish that there are different attitudes involved: if I miss my appointment, it does not really matter whether I overslept or had a flat tire; the result is the same. (Cf. Van Fraassen 2001, p167)

Maybe Van Fraassen is thinking of *good* reasons for belief, such that a theory’s explanatory or unifying power, elegance or simplicity does not count as a good reason. (Cf. his 2007, p345) This will not convince the realist who throws doubt on the differentia. The realist may consider some of these virtues to serve the epistemic aim and further our positive and fallible quest for truth, and the other virtues as providing no reason for acceptance

<sup>16</sup>This worry is found in Melchert (1985), Horwich (1991), Rosen (1994, p148–52), Psillos (1999, p200–4), Fine (2001), Teller (2001, p139–42), Ladyman (2007, p46–5) and Dicken (2010, ch4). Van Fraassen anticipates and responds in his (1980b, p87–9; 1983b, p165–8; 1985, p276–281; 1989, p189–94; 2001, p164–8; 2007, p341–6).

<sup>17</sup>These are only the terms used by Van Fraassen. We further find a cognitive versus non-cognitive commitment (Rosen 1994, p150; Psillos 1999, p201), assertion versus quasi-assertion (Rosen *Ibid.*), actual versus potential belief (Psillos *Ibid.*, p202), empirical versus non-empirical virtues (Ladyman 2007, p53) and pragmatic versus credential considerations (Dicken 2010, p158).

<sup>18</sup>Cf. Van Fraassen (1989, p192; 2003, p482), Horwich (1991, p2), Fine (2001, p109), Teller (2001, p126) and Dicken (2010, ch4, especially p150 fn2).

whatsoever. Van Fraassen's reliance on *good* reasons shows, at best, that the discrimination can be preached to the converted. At best: the empiricist should be allowed to believe in the EmpAd of some of our theories, but what *good* reasons can he offer to prefer one of two empirically rival theories that are both consistent with the evidence so far? The lack of compelling reasons for any specific ampliative move does not render such a commitment pragmatic.

The appeal to motives is also needlessly confusing. Van Fraassen holds that the virtues of a theory are those features that, in general, provide reasons for acceptance. (1983b, p166; 1985, p280) He contrasts confirmational and epistemic with informational and pragmatic virtues. Confirmational and epistemic virtues are features that make a theory more likely to be true. (1980b, p4; p89; 1983b, p166; 2007, p345) They purportedly provide the only proper reasons for belief. Other virtues are pragmatic or informational. "In so far as they go beyond consistency, empirical adequacy, and empirical strength, they [...] provide reasons to prefer the theory independently of questions of truth." (1980b, p88)

The *first* source of confusion is the implication that EmpAd provides a proper *reason* for belief. Consistency and simplicity are features that we can judge by solely considering the theory. EmpAd is not similarly recognizable; it cannot be read off of the theory and typically cannot be conclusively established. The belief in EmpAd is precisely what the epistemic commitment involved in theory acceptance amounts to. It cannot itself provide a reason for this belief. In another paper, Van Fraassen indeed states that the belief in EmpAd involves a pragmatically motivated leap of faith. To get us out of sceptical despair, we need courage: "a virtue which can enter when purely epistemic virtues are not far reaching enough." (2000, p273–4)

The *second*, related puzzling element is the equivocation of *confirmational* and *epistemic* virtues. Confirmational virtues are contrasted with informational virtues, but not all informational virtues are *ipso facto* non-epistemic. Empirical strength is an informational virtue par excellence, but it makes a theory less likely to be true. The prediction that tomorrow it will rain or not is likelier than the empirically stronger thesis that tomorrow it will rain. Yet, this should not render empirical strength a pragmatic virtue. If this is what Van Fraassen has in mind, it does not sit well with his allegiance to the Jamesian, two-fold construal of the epistemic aim: avoid falsehoods and believe truths. These desiderata cannot be jointly maximized, for they pull in opposite directions. To avoid falsehoods, we should believe nothing; to believe truths, we should believe any  $p$  and not- $p$ . Empirical strength detracts from the defensive component—it makes a theory more likely to be false—but it contributes to the other desideratum and serves the overall epistemic aim.<sup>19</sup>

---

<sup>19</sup>This is not to say that we should believe as much as possible. The desiderate pull in

Let us drop the contrast confirmational–informational and let empirical strength be among the epistemic virtues. EmpAd is not a reason for belief. The courageous faith is not dictated by a theory’s ‘purely epistemic virtues’, but we keep in mind that the faith *is* the epistemic commitment that comes aboard with theory acceptance. Teller retorts that this faith detracts from the epistemic aim. “If belief is to provide any sort of guide to life, we must stick out our necks and risk falsehood to gain informativeness. But now haven’t we already succumbed to desires tainted by the wish for something in addition to, and hence other than, just truth?” (2001, p140) But this is only correct if our desire for truth is spelled out as maximally avoiding falsehoods, which would be served best by believing nothing. Teller’s construal omits the second desideratum of the epistemic aim: to believe all truths.<sup>20</sup> To stick one’s neck out is to take a risk, but it is an *epistemic* risk.

There emerges a tripartition of theoretical virtues. At bottom, a theory should be internally consistent and consistent with the evidence. These are necessary conditions for acceptance. Acceptance is divided into an epistemic commitment, i.e. the belief in truth or EmpAd, and, for CE, a pragmatic commitment, based on such features as elegance and explanatory loveliness. Let the features of a theory that contribute to the epistemic aim be ‘epistemic virtues’. For CE these are consistency, consistency with the evidence and empirical strength. To commit epistemically, i.e. to belief in a theory’s EmpAd, is to take a wager. Although the choice between two not-yet-refuted but empirically non-equivalent theories may be based on pragmatic considerations, this does not make the belief in EmpAd a pragmatic commitment. (Nor does it make the motives any less pragmatic.) Motives for adopting one or the other commitment are irrelevant to the nature of this commitment.

With this conceptual confusion cleared away, we focus on the two faces of commitment—according to Horwich a distinction without a difference. He, as Melchert (1985) before him, sums up all the facets of the pragmatic commitment and then asks what more can be involved in belief. The answer, I take it, is simply this: when believing  $p$ , one thinks that  $p$  is true and one is not neutral regarding its truth. Evidence for this attitude can be found e.g. in verbal behaviour, such as the answer to the question, ‘do you sincerely believe  $p$ ?’ Horwich anticipates this response and claims that we can be confused about our own beliefs, such that the denial of a belief may be mistaken and ‘muddled by philosophical double-talk’. (1991, p4) On the other hand, he does consider ‘prompting certain utterances’ one of the key features that a psychologist would use in characterizing the nature of this state of mind. (*Ibid.*, p3)

Now, I am not sure about the functional characterisation, let alone the

---

*opposite* directions; we should aim for some reflexive equilibrium, without losing sight of doxastic prudence and responsibility.

<sup>20</sup>A comparable misunderstanding is found in Ladyman (2007, p52), Dicken (2010, p158) and, surprisingly, Van Fraassen (1983b, p168).

nature of the relevant psychological states of mind, but I think that the workaday distinction between belief and agnosticism is better off than, for instance, that between wanting, wishing, hoping, longing, desiring, et cetera. When I believe something, I may be willing to bet on it. I would put some money on the claim that if you were to open the drawer, you would see the letter. In the role of explainer, I could point out why we decided to not build the dangerous power plant, but putting my money on the literal truth—instead of the usefulness—of claims which I believe to be *counterfactual*, is *pointless*.

Still, these are merely some illustrative pointers in terms of behaviour. We are all acquainted with immersion in fiction. When discussing a theatre play, we temporarily step into the shoes of the protagonists, speculate whether Vladimir and Estragon shouldn't just have left, predict, explain, infer and deliberate. This outward behaviour satisfies Horwich's gestural definition of belief (*Ibid.*), but we are aware it is *just* pretence (and drop this involvement when leaving the theatre). The pragmatic commitment to a theory, by contrast, implies no negation, but agnosticism. It also involves the expectation of vindication: the confidence that the theoretical apparatus will remain fruitful in the future. Since we are phenomenologically familiar with *as-if*'s that we believe to be false, it is unproblematic to claim certain (pragmatic) virtues for a story about which we are epistemically neutral.

### 3.6 Conclusion

CE distinguishes sharply between *epistemic* and *pragmatic* commitment, where the former may be reserved to the empirical substructures of accepted science and the latter applies to any accepted residue structure. The two faces of commitment correspond to a distinction between the *empirical* realm, populated exclusively by the actual observable entities, and the *supra-empirical* realm, which, as far as science is concerned, might be a fable. Scientific supra-empirical stories are highly useful, possibly true, and maybe even indispensable, but *belief* in them is supererogatory.

The difference between the two kinds of commitment is commonsensical and straightforward: epistemic commitment and agnosticism are distinct propositional attitudes. Although the pragmatic commitment to theory and research programme is a substantial engagement, affecting the way one speaks of unobservables and counterfactuals and permeating explanation and experimental design, this commitment remains an epistemic *as-if*, i.e. the existence of those unobservables and counterfactual truth-makers is not asserted. In non-scientific contexts, we are all familiar with engagement-sans-belief, e.g. when we are immersed in fiction, as well as with the epistemic attitude of neutrality. Put immersion and epistemic neutrality together and you have a pragmatic commitment.

The realm of the actual consists of actual, individual entities—objects, events and processes—and encompasses past, present and future. We might say that individual objects exist, individual events happen and individual processes unfold. A theory is EmpAd if it has a model in which (the data models of) the actual entities within its domain, if observable, can be embedded. It does not cover objects that, intuitively, *could* have existed but do not, or events that *might* have occurred but did not. The belief in a theory's EmpAd, therefore, is not intended to involve belief in counterfactuals; those beliefs that have their workaday analogue in the 'what-if's so hotly debated by soccer fans. This example is telling, for it points towards an immediate problem with such stories. We cannot put them to the test; hence, there is no way to settle disputes on empirical grounds. For the empiricist, this means there is no way to settle scientifically disputes over counterfactuals at all.

To practice his policy, the empiricist has to assess the implications of his belief in a theory's EmpAd: he must make judgments on actuality. Suppose he accepts a theory  $T$  which entails that under certain general observable circumstances  $C$  some observable event  $E$  will occur, i.e.  $T, C \models E$ . Let  $c \Box \rightarrow e$  be a *specific* claim about the occurrence of event-token  $e$  conditional on the occurrence of the required circumstances  $c$ , such as the claim: 'had the cheap nuclear power plant been built, a meltdown would have occurred.' When confronted with this claim, the empiricist wants to judge the actuality of  $c$ . If he believes that  $c$  is defeated (he believes not- $c$ ), then  $e$  does not follow from the EmpAd of  $T$ , and the empiricist does not need to believe the conditional. If he believes  $c$ , he should believe that  $e$ , since  $T, c \models e$  and if not- $e$ , his theory is not EmpAd. If he is neutral regards  $c$ , then the belief in  $e$  is not implied. This is an expression and application of the outlook that general claims are only vulnerable to actual occurrences and that it is not needed to commit epistemically to their counterfactual adequacy.

This policy has consequences for the definition of observability. Van Fraassen's initial rough guide to judgements of observability is based on possible observations, i.e. conditionals of the form  $c \Box \rightarrow e$ , so if we unpack the policy it reads something like: as far as science is concerned, we only need to believe in claims about entities that are (i) actual and (ii) would be observed under possible, including non-actual, circumstances. The several options to maintain this explication are not appealing. We could apply condition (i) to condition (ii) and restrict our beliefs to observations under actual circumstances, but the resulting *manifestism* is untenable. We could commit *pragmatically* to non-actual observations, but then we are left with a 'mere' pragmatic policy. Or we commit *epistemically* to possible observations, rendering the policy, by its own lights, scientifically supererogatory. This needs not be disastrous *ipso facto*—compare Popper's criterion of falsifiability—but the belief in counterfactuals runs counter to the empiricist spirit. The stipulation that the policy is not self-applicable runs into the same episte-

mological trouble, as does the option of extra extending the policy so as to incorporate all possible observables.

Fortunately, the rough guide is no definition, although it may offer a helping hand in performing thought experiments to assess observability. The limits of the observable are a theory-independent question. Our revisable judgments of observability are based on current science. The theory of light tells us that entities are observable in virtue of an appropriate size and reflection spectrum. The assessments depend on how the entities of interest are modelled. Non-actual observation-events can be interpreted as a way of exploring the relevant features before the mind's eye. The empiricist does not have to believe these counterfactuals; acceptance of the theory of light suffices to guide his judgments of observability.

The semantic notion of EmpAd is relatively well understood. Hereby, I have thrown some light on how to work with the doxastic policy. As for the cheap nuclear power plant, we have, in truly Humean fashion, two reactions—both are correct. As a detached empiricist, we hold that no human knows what would have actually unfolded. As the immersed nuclear scientist, we are relieved that such a potential disaster is avoided.

# Conclusion

Constructive Empiricism (CE) is the view of science that holds that science is aimed at constructing theories that are adequate to the empirical, i.e. the actual observable. The empirical realm is populated by the phenomena: the actual observable objects, events and processes. In the light of this aim, theory acceptance implies as belief only that the theory is empirically adequate; the epistemic commitment to the supra-empirical (non-actual entities and unobservable entities) is scientifically supererogatory. In this thesis, I have argued that CE is the correct view of science. I present the conclusions in reverse order.

Chapter 3. CE holds that acceptance of a scientific theory need not involve the belief in the theory's adequacy to the supra-empirical realm, which is populated, if at all, by non-actual entities and/or by unobservable entities. Full acceptance of a theory implies as belief only that the theory is adequate to the actual observable entities. This is CE's doxastic policy: the answer to the venerable question: what to believe? There are two notable worries about this policy. *First*, science involves counterfactual claims and a view of science cannot ignore these and still make sense of science as we know it. *Second*, it has been argued that the policy is inconsistent because the empiricist needs beliefs about the supra-empirical to delineate the empirical.

Theory acceptance does not mirror the epistemic commitment to a crystal ball—the theory can be used to make predictions, but in isolation, it typically does not provide a complete description of all that happens in its domain from second to second. Although some theory  $T$  provides models of *types* of phenomena, acceptance of  $T$  does not involve the belief that  $T$  is adequate to non-actual *tokens* of a type. To figure out what is involved epistemically in acceptance, the empiricist should make judgments of the actual tokens. A theory cannot be tested against non-existent entities. CE relies, for its motivation, on this emphasis on vulnerability to experiential evidence. Yet, theory acceptance also involves a pragmatic commitment and the empiricist can commit pragmatically to the associated non-actual entities and counterfactual claims. As an immersed scientist, he could behave *as if* he believed that these claims are true. As an empiricist, he holds that such claims are scientifically supererogatory.

The worries about the inconsistency of the doxastic policy are ill-founded.

The limits of the realm of the actual are an objective matter of fact: they do not change with our judgments of it. The limits of the realm of the observable are equally factual once our epistemic community is specified. Whether an entity is observable depends on the actual features of the entity and its observability does not change with our judgments of it. To judge observability, the empiricist can make use of theory-informed, counterfactual thought experiments to help her in predicating the relevant features to the entity under consideration. The belief that some theory is empirically adequate is the belief that the phenomena within its domain are saved by the theory, in which case any alleged non-empirical entity in its domain is either unobservable or counterfactual. The empiricist can, thus, base his judgments of the extent of the epistemic commitment involved in theory acceptance on theories he accepts.

Chapter 2. The empiricist can, thus, consistently judge observability and restrict his beliefs to the empirical, but why should he? It has been argued that we need to believe in science's adequacy to the unobservable in order to explain science's empirical and technological success. If this success does not spring from the approximate or partial adequacy to the unobservable, it is inexplicable and *that* is unacceptable. This is the no-miracles intuition and the proposed inference to realism is an abduction: an inference from some observation to the best explanation of this observation. Abductions rely on the supposition that explanatory virtuousness (loveliness) makes a theory likely. Regrettably for realism, we have no reasons to believe this supposition.

One route to argue for realism is to hold that we are forced to accept the best explanation, whether explanatory virtuousness is an epistemic virtue or not. This argument fails: we cannot be forced to infer the best explanation if the explanation is simply not good enough. Moreover, there is a point at which the demand for explanation must be rejected. Just as the realist is satisfied with an explanation by truth, the empiricist is satisfied with an explanation by empirical adequacy, and neither of them can be forced to explain further. The perceived correlation between empirical success and loveliness does not establish that loveliness is a symptom of truth; the correlation admits of rival interpretations. For instance, scientists tweak successful theories to make them lovelier, or the scientific standard of loveliness is informed by the form of successful theories. In neither case is loveliness a reliable guide to likeliness. Finally, some often mentioned explications of loveliness, e.g. precision or broad scope, evidently make a theory less likely.

The aim of information-oriented activity is two-fold: believe all truths and believe no falsehood. Theoretical features such as precision and broad scope make a theory more informative and therefore more likely to be false. Yet, if we wanted to avoid believing falsehoods at all costs, we had better restrict our beliefs to tautologies. Informativeness is a virtue in the light of the other epistemic desideratum: believe all truths. As such, the inference

of an explanation promotes the epistemic aim just as the adoption of any belief does. Yet, it is unclear why loveliness is a specific symptom of truth. By contrast, it is evident how aesthetic judgments guide theory choice and how the strive for loveliness helps to keep our ontologies simple and comprehensible. At a minimum, we conclude that we are not compelled to infer the loveliest theory. Yet, somewhat more ambitious, we currently have no more reasons to think of loveliness as an epistemic virtue than we have to think of e.g. explanatory ugliness as an epistemic virtue. Loveliness is plausibly a pragmatic virtue.

Chapter 1. CE's doxastic policy is consistent and we are not compelled to believe in non-actuals or unobservables. Additionally, CE's is the correct aim of science; theory acceptance should not involve more or less beliefs than the belief in the theory's empirical adequacy.

Theory acceptance should not involve more, i.e. beliefs about the supra-empirical, because it is plausible that these beliefs are evidentially invulnerable and the adoption of invulnerable beliefs is epistemically reckless. An ampliative belief is invulnerable if there is no evidence that could conflict with it, and the adoption of such a belief is reckless because its possible falsity cannot be exposed. The only source of evidence that is widely accepted in science is experience. The most promising alternative source of evidence, explanatory loveliness, has been found wanting. Thus, we may cautiously conjecture that the only scientifically permissible evidence is empirical. If this conjecture is correct, then the adoption of supra-empirical beliefs is not merely scientifically supererogatory, but also irresponsible.

The acceptance of a theory may involve less than the belief that the theory is empirically adequate, e.g. if the acceptance is partial or qualified. From an empiricist point of view, a scientist may—is rationally allowed to—be an inductive skeptic. However, inductive skepticism is unable to make sense of scientific activity, like experiment and prediction, and therefore fails as a general view of science. Several proposals to portray science as aimed at something less ambitious than empirical adequacy fall prey to the same problem. Manifestism, for instance, holds that science aims at adequacy to our actual observations. In a sense, this is trivially correct, but since we are unable to make principled and stable judgments of the distinction between observed and unobserved observables (past, present and future), this aim reduces, in practice, to the strive for empirical adequacy. This is why CE is a consistent, prudent, sensible and well-motivated view of science and, given the epistemological considerations pertaining to our 'epistemic reach', it is, plausibly, the correct view.

By way of *encore*, I suggest a few venues for future research. First of all, empiricism is in dire need of a clear and stable explication of the Janus-faced character of experience. As Radder notes (2006, p11–2), it is rather peculiar that the central notion of empiricism is so poorly understood. Nagel (2000) and Ladyman (2007) argue forcefully that no stable notion of experience is to

be had. If they are correct, the whole empiricist project falls to pieces. Now, I have argued that the problem of arriving at a non-foundational yet ‘worldly’ characterization of experience is a problem shared by all science-admiring philosophical stances/epistemologies. Still, such a notion of our main—and possibly only—source of scientific evidence would greatly benefit empiricism and may even aid it in its identity crisis: it could help with the formulation of a metaphysically innocuous version of the primacy of experience thesis.

Recently, new ‘algorithmic’ accounts of rationality spring from post-Kuhnian soil. These accounts, such as Williamson (2010), aim to put rationally compelling constraints on the choice of our initial credences, with an appeal to maximum entropy. The sketch of our responsibility-based account of rationality tends towards the ‘axiological’ pole, providing room for rational disagreement in the light of equal evidence and a subjective balancing of one’s values. This ‘ethics of beliefs’-approach does not sit well with ampliative recipes. The broadly voluntarist line, pleading for doxastic freedom, can be strengthened by formulating a general, encompassing criticism on ampliative algorithms.

Another direction that may prove fruitful is this: the pragmatic commitment to certain counterfactuals in the accepted models may partially accommodate causal talk. To make this connection, I take a few rather bold steps, so bear with me. Our causal talk is pervaded by a counterfactual dimension, roughly:  $c$  causes  $e$ , since  $\text{not-}c \square \rightarrow \text{not-}e$ . It also has a context-sensitive dimension, although this dimension is commonly dissected at the level of causal *explanation*, where we pick out the ‘salient’ nodes in the ‘wordly’ causal networks. (Cf. Van Fraassen 1980, p124; Lipton 2004, p144) Yet, if we look at two of the most elaborate counterfactual analyses of causation, namely Lewis’s (2000) and Woodward’s (2003), this context-sensitivity runs much deeper; all the way to these ‘objective’ causal networks.

Lewis’s analysis of causation in terms of influence allows us to decide, based on our ‘mood’, which ‘alterations’ of an event are relevantly close. (Lewis 2000, p197) This means that anything can be considered a cause of a subsequent event if we stretch the relevant proximity. The same problem plagues Woodward’s context-sensitive concept of ‘serious possibilities’. (Woodward 2003, p90) If we stretch this set sufficiently, any event is trivially caused by every prior event. If a ball is thrown at a window, but a fielder catches it, we say the fielder prevented the shattering of the window, i.e. caused it to not shatter. Yet, if there is a solid wall behind the catcher, we judge his catch irrelevant. Neither of the analyses is able to capture this judgment if we add the possibility of the ball magically flying through the solid wall. Enriched with this ‘alteration’ it is evaluated as genuine causation. All this needs to be argued for, but it should suffice in order to make clear how the pragmatic commitment is connected.

The context for subjunctive conditionals can be settled to some extent by considering the accepted models of physics. If there is no model for

the magical penetration of the wall, the causal conditional should not be accepted. Hence, the commitment to a theory puts some pragmatic constraints on the causal claims one is prepared to avow. There is still a lot of causation to choose from: even ‘within’ these models, we make choices based on salience, contrasts, relevance, normality, and so on. However, it is a good start to show that causal claims refer to—and are pragmatically constrained by—the accepted scientific models. Van Fraassen has presented anti-realist interpretations and eliminations of necessity, essences, explanation, laws and chance, but not of causation. (Although his account of explanation is based on causation, he left the latter concept unanalysed.) Empiricist reconstructions of scientific modal discourse promote the empiricist project, because one of the charges against CE is that it cannot make sense of actual scientific discourse.

# References

- Alspecter-Kelly, M. (2001). Should the Empiricist Be a Constructive Empiricist? *Philosophy of Science*, 413–431.
- Alspecter-Kelly, M. (2004). Seeing the Unobservable: Van Fraassen and the Limits of Experience. *Synthese* 140(3), 331–353.
- Alspecter-Kelly, M. (2006). Constructive Empiricism and Epistemic Modesty: Response to Van Fraassen and Monton. *Erkenntnis* 64(3), 371–379.
- Alspecter-Kelly, M. (2012). Constructive Empiricism Revisited. *Metascience* 21, 187–191.
- Barnes, E. (1995). Inference to the Loveliest Explanation. *Synthese* 103(2), 251–277.
- Bartha, P. (2004). Countable Additivity and the De Finetti Lottery. *The British Journal for the Philosophy of Science* 55(2), 301–321.
- Bayes, T. (1764). An Essay towards Solving a Problem in the Doctrine of Chances. *Philosophical Transactions* 53, 370–418.
- Berg-Hildebrand, A. and C. Suhm (2006). The Hardships of an Empiricist. In A. Berg-Hildebrand and C. Suhm (Eds.), *Bas C. Van Fraassen: The Fortunes of Empiricism*, pp. 57–67. Ontos Verlag.
- Bovens, L. and S. Hartmann (2004). *Bayesian Epistemology*. Oxford University Press.
- Boyd, R. N. (1983). On the Current Status of the Issue of Scientific Realism. In C. G. Hempel, H. Putnam, and W. K. Essler (Eds.), *Methodology, Epistemology, and Philosophy of Science*, pp. 45–90. Springer.
- Buekens, F. A. I. and F. A. Muller (2012). Intentionality Versus Constructive Empiricism. *Erkenntnis* 76, 91–100.
- Carnap, R. (1945). The Two Concepts of Probability: The Problem of Probability. *Philosophy and Phenomenological Research* 5(4), 513–532.
- Cartwright, N. (1983). *How the Laws of Physics Lie*. Cambridge University Press.

- Cartwright, N. (1999). *The Dappled World: A Study of the Boundaries of Science*. Cambridge University Press.
- Cartwright, N. (2007). Why Be Hanged for Even a Lamb? In B. Monton (Ed.), *Images of Empiricism: Essays on Science and Stances, with a Reply from Bas C. Van Fraassen*, pp. 32–45. Oxford; New York: Oxford University Press.
- Christensen, D. (1991). Clever Bookies and Coherent Beliefs. *The Philosophical Review* 100(2), 229–247.
- Churchland, P. M. (1985). The Ontological Status of Observables: In Praise of the Superempirical Virtues. In P. M. Churchland and C. A. Hooker (Eds.), *Images of Science: Essays on Realism and Empiricism, with a Reply from Bas C. Van Fraassen*, pp. 35–47. Chicago and London: The University of Chicago Press.
- Clifford, W. K. (1879). *Lectures and Essays by William Kingdon Clifford, F.R.S.*, Volume 2. London: Macmillan and Co.
- De Finetti, B. ([1937], 1992). Foresight: Its Logical Laws, Its Subjective Sources. In *Breakthroughs in Statistics*, pp. 134–174. Springer.
- Dicken, P. (2007). Constructive Empiricism and the Metaphysics of Modality. *The British Journal for the Philosophy of Science* 58(3), 605–612.
- Dicken, P. (2009a). Constructive Empiricism and the Vices of Voluntarism. *International Journal of Philosophical Studies* 17(2), 189–201.
- Dicken, P. (2009b). On the Syntax and Semantics of Observability: A Reply to Muller and Van Fraassen. *Analysis* 69(1), 38–42.
- Dicken, P. (2010). *Constructive Empiricism: Epistemology and the Philosophy of Science*. Houndmills, Basingstoke, Hampshire; New York: Palgrave Macmillan.
- Dicken, P. and P. Lipton (2006). What Can Bas Believe? Musgrave and Van Fraassen on Observability. *Analysis* 66(3), 226–233.
- Earman, J. (1992). *Bayes or Bust?* Cambridge: The MIT Press.
- Eddington, A. (1928). *The Nature of the Physical World*. New York: Macmillan.
- Fine, A. (1991). The Natural Ontological Attitude. *The Philosophy of Science*, 261–77.
- Fine, A. (2001). The Scientific Image Twenty Years Later. *Philosophical Studies* 106(1), 107–122.
- Fitelson, B. (2001). *Studies in Bayesian Confirmation Theory*. Ph. D. thesis, University of Wisconsin.
- Foley, R. (1990). Skepticism and Rationality. In M. Roth and G. Ross (Eds.), *Doubting*, pp. 69–81. Springer.

- Foley, R. (1992). Being Knowingly Incoherent. *Noûs* 26(2), 181–203.
- Foley, R. (2001). *Intellectual Trust in Oneself and Others*. Cambridge University Press.
- Foley, R. (2005). Is Justified Belief Responsible Belief? In M. Steup and E. Sosa (Eds.), *Contemporary Debates in Epistemology*, Contemporary Debates in Philosophy, pp. 313–326. Malden: Blackwell Publishing.
- Gaifman, H. (1988). A Theory of Higher Order Probabilities. In *Causation, Chance and Credence*, pp. 191–219. Springer.
- Ghins, M. (2002). Putnam’s No-Miracle Argument: A Critique. In *Recent Themes in the Philosophy of Science*, pp. 121–137. Springer.
- Gijssbers, V. A. (2011). *Explanation and Determination*. Philosophy Department, Faculty of the Humanities, Leiden University.
- Hacking, I. (1985). Do We See through a Microscope? In P. M. Churchland and C. A. Hooker (Eds.), *Images of Science: Essays on Realism and Empiricism, with a Reply from Bas C. Van Fraassen*, pp. 132–152. University of Chicago Press.
- Hanna, J. F. (2004). Contra Ladyman: What Really Is Right with Constructive Empiricism. *The British Journal for the Philosophy of Science* 55(4), 767–777.
- Harman, G. H. (1965). The Inference to the Best Explanation. *The Philosophical Review* 74(1), 88–95.
- Hájek, A. A Puzzle About Degree of Belief. (*unpublished*).
- Hájek, A. (2003). What Conditional Probability Could not Be. *Synthese* 137(3), 273–323.
- Hájek, A. and M. Smithson (2011). Rationality and Indeterminate Probabilities. *Synthese* 187, 33–48.
- Hofer, C. (2012). Calibration: Being in Tune with Frequencies. *Dialectica* 66(3), 435–452.
- Horwich, P. (1991). On the Nature and Norms of Theoretical Commitment. *Philosophy of Science*, 1–14.
- Howson, C. and P. Urbach (2006). *Scientific Reasoning: The Bayesian Approach* (3 ed.). Chicago: Open Court.
- Iranzo, V. (2008). Bayesianism and Inference to the Best Explanation. *Theoria. Revista de Teoría, Historia y Fundamentos de la Ciencia* 23(1), 89–106.
- James, W. (1896). The Will to Believe. *The New World* V, 327–347.
- Jauernig, A. (2007). Must Empiricism Be a Stance, and Could it Be One? How to Be an Empiricist and a Philosopher at the Same Time. In B.

- Monton (Ed.), *Images of Empiricism: Essays on Science and Stances, with a Reply from Bas C. Van Fraassen*, pp. 271–318. Oxford: Oxford University Press.
- Jeffrey, R. (1965). *The Logic of Decision*. New York: McGraw-Hill.
- Jeffrey, R. (1983). Bayesianism with a Human Face. *Testing Scientific Theories, Minnesota Studies in the Philosophy of Science 10*, 133–56.
- Jeffrey, R. (2004). *Subjective Probability: The Real Thing*. Cambridge University Press.
- Keynes, J. (1921). *A Treatise on Probability*. London: Macmillan.
- Klukhuhn, A. (2008). *Alle Mensen Heten Janus: Het Verbond tussen Filosofie, Wetenschap, Kunst en Godsdienst*. Amsterdam: Bert Bakker.
- Kolmogorov, A. N. (1956). *Foundations of the Theory of Probability*. (2 ed.). New York: Chelsea Publishing Company.
- Kuhn, T. (1962). *The Structure of Scientific Revolutions*. University of Chicago Press.
- Kvanvig, J. L. (1994). A Critique of Van Fraassen's Voluntaristic Epistemology. *Synthese 98*(2), 325–348.
- Ladyman, J. (2000). What's Really Wrong with Constructive Empiricism? Van Fraassen and the Metaphysics of Modality. *The British Journal for the Philosophy of Science 51*(4), 837–856.
- Ladyman, J. (2004a). Constructive Empiricism and Modal Metaphysics: A Reply to Monton and Van Fraassen. *The British Journal for the Philosophy of Science 55*(4), 755–765.
- Ladyman, J. (2004b). Discussion—Empiricism Versus Metaphysics. *Philosophical Studies 121*(2), 133–145.
- Ladyman, J. (2007). The Epistemology of Constructive Empiricism. In B. Monton (Ed.), *Images of Empiricism: Essays on Science and Stances, with a Reply from Bas C. Van Fraassen*, pp. 46–61. Oxford: Oxford University Press.
- Ladyman, J. (2011). The Scientistic Stance: The Empirical and Materialist Stances Reconciled. *Synthese 178*, 87–98.
- Ladyman, J., I. Douven, L. Horsten, and B. C. Van Fraassen (1997). A Defence of Van Fraassen's Critique of Abductive Inference: Reply to Psillos. *The Philosophical Quarterly 47*(188), 305–321.
- Ladyman, J., D. Ross, D. Spurrett, and J. G. Collier (2007). *Every Thing Must Go: Metaphysics Naturalized*, Volume 61. Oxford University Press Oxford.

- Laudan, L. (1981). A Confutation of Convergent Realism. *Philosophy of Science*, 19–49.
- Leeds, S. (1994). Constructive Empiricism. *Synthese* 101(2), 187–221.
- Leplin, J. (1982). The Historical Objection to Scientific Realism. In *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, pp. 88–97.
- Levi, I. (1974). On Indeterminate Probabilities. *The Journal of Philosophy* 71(13), 391–418.
- Lewis, D. K. ([1980] 1986a). A Subjectivist's Guide to Objective Chance. In *Philosophical Papers*, Volume 2, pp. 83–113. Oxford University Press.
- Lewis, D. K. (1986b). Events. In *Philosophical Papers*, Volume 2, pp. 241–269. Oxford University Press.
- Lewis, D. K. (1994). Humean Supervenience Debugged. *Mind* 103(412), 473–490.
- Lewis, D. K. (2000). Causation as Influence. *The Journal of Philosophy* 97(4), 182–197.
- Lipton, P. (2004). *Inference to the Best Explanation* (2 ed.). Routledge.
- Lipton, P. (2005). Wouldn't it be Lovely: Explanation and Scientific Realism. *Metascience* 14(3), 353–361.
- Maxwell, G. (1962). The Ontological Status of Theoretical Entities. In H. Feigl and G. Maxwell (Eds.), *Scientific Explanation, Space, and Time*, Volume 3 of *Minnesota Studies in the Philosophy of Science*. Minneapolis: University of Minnesota Press.
- McAllister, J. W. (1996). *Beauty and Revolution in Science*. Ithaca, London: Cornell University Press.
- McMullin, E. (1984). A Case for Scientific Realism. In J. Leplin (Ed.), *Scientific Realism*, pp. 8–40.
- Melchert, N. (1985). Why Constructive Empiricism Collapses into Scientific Realism. *Australasian Journal of Philosophy* 63(2), 213–215.
- Monton, B. and B. C. Van Fraassen (2003). Constructive Empiricism and Modal Nominalism. *The British Journal for the Philosophy of Science* 54(3), 405–422.
- Muller, F. A. (2004). Can a Constructive Empiricist Adopt the Concept of Observability? *Philosophy of Science* 71, 80–97.
- Muller, F. A. (2005). The Deep Black Sea: Observability and Modality Afloat. *The British Journal for the Philosophy of Science* 56, 61–99.
- Muller, F. A. and B. C. Van Fraassen (2008). How to Talk about Unobservables. *Analysis* 68(299), 197–205.

- Musgrave, A. (1982). Review: Constructive Empiricism Versus Scientific Realism. *The Philosophical Quarterly* 32(128), 262–271.
- Musgrave, A. (1985). Realism versus Constructive Empiricism. In P. M. Churchland and C. A. Hooker (Eds.), *Images of Science: Essays on Realism and Empiricism, with a Reply from Bas C. Van Fraassen*, Volume 1985, pp. 197–221.
- Musgrave, A. (1988). The Ultimate Argument for Scientific Realism. In *Relativism and Realism in Science*, pp. 229–252. Springer.
- Musgrave, A. (2007). The Miracle Argument for Scientific Realism. *The Rutherford Journal: The New Zealand Journal for the History and Philosophy of Science and Technology* 2.
- Nagel, J. (2000). The Empiricist Conception of Experience. *Philosophy* 75(03), 345–376.
- Niiniluoto, I. (1999). Defending Abduction. *Philosophy of Science*, S436–S451.
- Okasha, S. (2000). Van Fraassen's Critique of Inference to the Best Explanation. *Studies In History and Philosophy of Science Part A* 31(4), 691–710.
- Popper, K. ([1934, 1958] 2009). *The Logic of Scientific Discovery*. London; New York: Routledge.
- Psillos, S. (1996). On Van Fraassen's Critique of Abductive Reasoning. *The Philosophical Quarterly* 46(182), 31–47.
- Psillos, S. (1997). How Not to Defend Constructive Empiricism: A Rejoinder. *The Philosophical Quarterly* 47(188), 369–372.
- Psillos, S. (1999). *Scientific Realism: How Science Tracks Truth*. London; New York: Routledge.
- Psillos, S. (2007a). The Fine Structure of Inference to the Best Explanation. *Philosophy and Phenomenological Research* 74(2), 441–448.
- Psillos, S. (2007b). Putting a Bridle on Irrationality: An Appraisal of van Fraassen's New Epistemology. In B. Monton (Ed.), *Images of Empiricism: Essays on Science and Stances, with a Reply from Bas C. Van Fraassen*, pp. 134–164. Oxford University Press.
- Psillos, S. (2009). *Knowing the Structure of Nature: Essays on Realism and Explanation*. Palgrave Macmillan.
- Psillos, S. (2011a). Choosing the Realist Framework. *Synthese* 180, 301–316.
- Psillos, S. (2011b). The Scope and Limits of the No Miracles Argument. In D. Dieks (Ed.), *Explanation, Prediction, and Confirmation*, pp. 23–35. Springer.

- Putnam, H. (1975). *Mathematics, Matter and Method: Philosophical Papers, Volume I*. Cambridge University Press.
- Quine, W. V. (1951). Two Dogmas of Empiricism. *The Philosophical Review* 60(1), 20–43.
- Radder, H. (2006). *The World Observed / The World Conceived*. University of Pittsburgh Press.
- Railton, P. (1989). Explanation and Metaphysical Controversy. In P. Kitcher and W. C. Salmon (Eds.), *Scientific Explanation, Volume 8 of Minnesota Studies in the Philosophy of Science*, pp. 220–252.
- Ramsey, F. P. ([1926] 1931). Truth and Probability. *The Foundations of Mathematics and Other Logical Essays*, 156–198.
- Rosen, G. (1994). What Is Constructive Empiricism? *Philosophical Studies* 74(2), 143–178.
- Salmon, W. C. (1981). Rational Prediction. *The British Journal for the Philosophy of Science* 32(2), 115–125.
- Samuels, R. and S. Stich (2004). Rationality and Psychology. In A. R. Mele and P. Rawling (Eds.), *The Oxford Handbook of Rationality*, pp. 279–300. Oxford University Press.
- Schurz, G. Ostensive Learnability as a Test Criterion for Theory-Neutral Observation Concepts. (*unpublished*).
- Sellars, W. (1962). Philosophy and the Scientific Image of Man. In R. Colodny (Ed.), *Frontiers of Science and Philosophy*. University of Pittsburgh Press.
- Strevens, M. (1999). Objective Probability as a Guide to the World. *Philosophical Studies* 95(3), 243–275.
- Teller, P. (1973). Conditionalization and Observation. *Synthese* 26(2), 218–258.
- Teller, P. (2001). Whither Constructive Empiricism? *Philosophical Studies* 106(1), 123–150.
- Teller, P. (2011). Learning to Live with Voluntarism. *Synthese* 178, 49–66.
- Van Dyck, M. (2007). Constructive Empiricism and the Argument from Underdetermination. In B. Monton (Ed.), *Images of Empiricism: Essays on Science and Stances, with a Reply from Bas C. Van Fraassen*, pp. 11–31. Oxford: Oxford University Press.
- Van Fraassen, B. C. (1980a). Rational Belief and Probability Kinematics. *Philosophy of Science*, 165–187.
- Van Fraassen, B. C. (1980b). *The Scientific Image*. Oxford University Press.

- Van Fraassen, B. C. (1983a). Calibration: A Frequency Justification for Personal Probability. In *Physics, Philosophy and Psychoanalysis*, pp. 295–319. Springer.
- Van Fraassen, B. C. (1983b). Glymour on Evidence and Explanation. *Minnesota Studies in Philosophy of Science* 10, 165–176.
- Van Fraassen, B. C. (1984). Belief and the Will. *The Journal of Philosophy* 81(5), 235–256.
- Van Fraassen, B. C. (1985). Empiricism in the Philosophy of Science. In P. M. Churchland and C. A. Hooker (Eds.), *Images of Science: Essays on Realism and Empiricism, with a Reply from Bas C. Van Fraassen*, pp. 245–308.
- Van Fraassen, B. C. (1989). *Laws and Symmetry*. Oxford: Clarendon Press.
- Van Fraassen, B. C. (1990). Figures in a Probability Landscape. In *Truth or Consequences*, pp. 345–356. Springer.
- Van Fraassen, B. C. (1992). From Vicious Circle to Infinite Regress, and Back Again. In *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, pp. 6–29.
- Van Fraassen, B. C. (1994a). Against Transcendental Empiricism. In *The Question of Hermeneutics*, pp. 309–335. Springer.
- Van Fraassen, B. C. (1994b). Gideon Rosen on Constructive Empiricism. *Philosophical Studies* 74(2), 179–192.
- Van Fraassen, B. C. (1995a). Against Naturalized Epistemology. *On Quine*, Cambridge University Press, Cambridge, 68–88.
- Van Fraassen, B. C. (1995b). Belief and the Problem of Ulysses and the Sirens. *Philosophical Studies* 77(1), 7–37.
- Van Fraassen, B. C. (1999). Conditionalization, a New Argument for. *Topoi* 18(2), 93–96.
- Van Fraassen, B. C. (2000). The False Hopes of Traditional Epistemology. *Philosophical and Phenomenological Research*, 253–280.
- Van Fraassen, B. C. (2001). Constructive Empiricism Now. *Philosophical Studies* 106, 151–170.
- Van Fraassen, B. C. (2002). *The Empirical Stance*. Yale University Press.
- Van Fraassen, B. C. (2003). On McMullin's Appreciation of Realism Concerning the Sciences. *Philosophy of Science* 70(3), 479–492.
- Van Fraassen, B. C. (2004). Replies to Discussion on the Empirical Stance. *Philosophical Studies* 121(2), 171–192.

- Van Fraassen, B. C. (2005). The Day of the Dolphins: Puzzling over Epistemic Partnership. In K. Peacock and A. Irvine (Eds.), *Mistakes of Reason: Essays in Honour of John Woods*. Toronto: University of Toronto Press.
- Van Fraassen, B. C. (2006). Replies. In *Bas C. Van Fraassen: The Fortunes of Empiricism*, pp. 125–171. Ontos Verlag.
- Van Fraassen, B. C. (2007). From a View of Science to a New Empiricism. In B. Monton (Ed.), *Images of Empiricism: Essays on Science and Stances, with a Reply from Bas C. Van Fraassen*, pp. 337–383.
- Van Fraassen, B. C. (2008). *Scientific Representation: Paradoxes of Perspective*. Oxford University Press.
- Van Fraassen, B. C. (2011). On Stance and Rationality. *Synthese* 178(1), 155–169.
- Votsis, I. (2011). Structural Realism: Continuity and Its Limits. In P. Bokulich and A. Bokulich (Eds.), *Scientific Structuralism*. Springer.
- Weisberg, J. (2008). Locating IBE in the Bayesian Framework. *Synthese* 167(1), 125–143.
- Weisberg, J. (2009). Varieties of Bayesianism. In D. M. Gabbay, S. Hartmann, and J. Woods (Eds.), *Handbook of the History of Logic: Inductive Logic*, Volume 10, pp. 477–551.
- Williamson, J. (1999). Countable Additivity and Subjective Probability. *The British Journal for the Philosophy of Science* 50(3), 401–416.
- Williamson, J. (2010). *In Defence of Objective Bayesianism*. Oxford; New York: Oxford University Press.
- Wittgenstein, L. ([1953] 2009). *Philosophical Investigations*. Malden: Blackwell.
- Woodward, J. (2005). *Making Things Happen: A Theory of Causal Explanation*. Oxford University Press.
- Worrall, J. (1989). Structural Realism: The Best of Both Worlds? *Dialectica* 43(1-2), 99–124.
- Worrall, J. (2011). The No Miracles Intuition and the No Miracles Argument. In D. Dieks (Ed.), *Explanation, Prediction, and Confirmation*, pp. 11–21. Springer.