

UNIVERSIDADE ESTADUAL DE CAMPINAS
INSTITUTO DE FILOSOFIA E CIÊNCIAS HUMANAS

Daniel Credico de Coimbra

**METAPHYSICAL EXPLANATION AND
THE INFERENCE TO THE BEST EXPLANATION**

CAMPINAS

2018

Daniel Credico de Coimbra

Metaphysical Explanation and the Inference to the Best Explanation

Monografia apresentada ao Instituto
de Filosofia e Ciências Humanas da
Universidade Estadual de Campinas

Orientador: Prof. Dr. Marco Antonio
Caron Ruffino

CAMPINAS

2018

Ata Filosofia/IFCH - 02/2018

ATA DE MONOGRAFIA

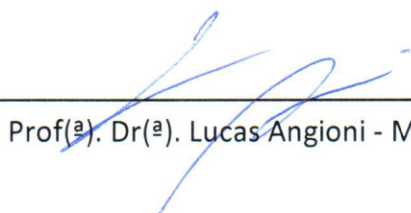
Curso de Filosofia
Instituto de Filosofia e Ciências Humanas
Universidade Estadual de Campinas
Disciplina HG880 - Monografia II

Em 12 de dezembro de 2018, no Instituto de Filosofia e Ciências Humanas da Universidade Estadual de Campinas, reuniram-se os professores Lucas Angioni, do Departamento de Filosofia/IFCH/UNICAMP e Pedro Merluzzi, do Departamento de CLE/IFCH/UNICAMP, membros da banca, sob a responsabilidade do Prof. Dr. Marco Antonio Caron Ruffino, do Departamento de Filosofia/IFCH/UNICAMP, orientador(a) da monografia intitulada "Metaphysical Explanation and the Inference to the Best Explanation", do(a) aluno(a) Daniel Credico de Coimbra.

O(A) aluno(a) foi aprovado(a) com nota: 10,0



Prof^(a). Dr^(a). Marco Antonio Caron Ruffino – Orientador(a)



Prof^(a). Dr^(a). Lucas Angioni - Membro da Banca



Prof^(a). Dr^(a). Pedro Merluzzi - Membro da Banca

*Dedicated to my dear mom.
You'd be surprised at your dumb teenage son
having grown quite fond of academics!
(I've grown a funny beard too.)*

ACKNOWLEDGMENTS

So-called *scientific philosophy* is a funny business. It aims at truth and proceeds rigorously, yet something seems amiss. Somehow the results seem couched on no more than bare judgment (“intuition”), and rigorous philosophy boils down to the meticulous management of these judgments. I would perhaps not have grown so concerned over this methodology if it were not for the bothersome insistence of a scientifically-minded friend of mine, Alírio Moura. To him, I owe many intellectual debts. Those concerns have led me to work on a methodology of theory choice in philosophy grounded in a theory’s explanatory goodness rather than its intuitive content (even if judgments of explanatory goodness are ultimately grounded in intellectual intuition). One of the results of this work is the present monograph.

Rationality, in turn, is a treacherous enterprise. Humans cannot reliably track the truth by themselves; epistemic rationality occurs at the community level. Nevertheless, too much deference to popular academic opinion is intellectually stifling in more than one way. Somehow, two groups I’m a part of have managed to strike an interesting equilibrium between individual intellectual autonomy and respect for peer opinion. They are the ALEPH undergraduate study group and Marco Ruffino’s research group in epistemology and philosophy of language, both at the University of Campinas. I thank its members for giving me enough elbow room to construct my own perspective, but enough pressure to prevent my brain from swirling out of my head in self-indulging speculation.

Among the many who have corrected my theoretical excesses or commented on my work are Felipe Albarelli, Lucas Angioni, Gustavo de Azevedo, Davi Bastos, Gustavo Bertolino, Emiliano Boccardi, Walter Carnielli, Vinícius Carvalho, Iana Cavalcanti, Lauro Edison, Francesco Ferrari, Gregory Gaboardi, David Horst, Monique Hulshof, Rony Marques, Gabriel Maruchi, Filipe Martone, Pedro Merluzzi, Ivore Mira, Valdenor Monteiro, Alexandre Reggiolli, Laura Rifo, Marco Ruffino, Nicola Salvatore, Renato Semaniuc, Victor Sholl, Leonardo Soutello, Hélio Steven, Joshua Thorpe, Giorgio Venturi, Rafael Viana, Hiro Watanabe, and Júlio Zampietro. Thanks, folks. Life is wonderful when foundational topics, no matter how practically inert, can be discussed with such an ensemble of active, bright minds. I could only wish for more natural, cognitive, and formal scientists in my peer group!

Since the life of the mind takes such a long time to return any financial benefit, my father, Márcio, deserves special gratitude for taking care of things while the world moved on around my study room, and for the emotional support he and my family have provided all this time.

ABSTRACT

Inference to the Best Explanation, roughly put, appeals to the explanatory power of a theory or hypothesis (relative to some data set) as constituting epistemic justification for it. Inference to the Best Explanation (henceforth IBE) is a tool widely employed among all reasoners alike, from the empirical sciences to ordinary life. Philosophical discussions do not differ in the usualness of explanatory appeals of this kind during serious argument. Often enough, the appeal is dialectically blocked, as many of our epistemic peers in philosophy offer reasons to be skeptical of IBE. Our aim with this monograph is to assess one worry that have been raised about this mode of inference: That explanatory power is not truth-conducive. We begin by discussing general features of inferences and then formulating IBE in detail. Afterward, we explicate and apply a canonical understanding of what an explanation is. This will lead to a certain understanding of explanatory power. We undergo a case study to defend the thesis that this kind of explanatory power is indeed epistemically irrelevant – unless, perhaps, when combined with other theoretical virtues. Our conclusion is that the measure what explanations are *best* requires taking other theoretical virtues into account, such as simplicity and unification. In this case, a complete assessment of IBE requires examining *if, when, and how* these alleged theoretical virtues are indeed truth-conducive.

Key-words: Philosophy of science; theory choice; inference to the best explanation

RESUMO

A inferência à melhor explicação, grosso modo, apela ao poder explanatório de uma teoria ou hipótese (relativamente a algum conjunto de dados) como constituindo justificção epistêmica para ela. A inferência à melhor explicação (doravante “IBE”, do inglês *inference to the best explanation*) é uma ferramenta amplamente empregada por todos os raciocinadores, das ciências empíricas à vida comum. Discussões filosóficas não diferem na habitualidade de apelos explanatórios deste tipo durante argumentações sérias. Frequentemente, o apelo é dialeticamente bloqueado, dado que muitos de nossos pares epistêmicos em filosofia oferecem razões para que se seja cético quanto à IBE. Nosso objetivo com esta monografia é avaliar uma das preocupações que foram levantadas quanto a este modo de inferência: Que o poder explanatório não é verocondutor. Nós começamos discutindo características gerais de inferências e então formulando a IBE em detalhes. Depois, oferecemos uma visão canônica sobre o que uma explicação é. Isto levará a uma certa compreensão de poder explanatório. Analisaremos um estudo de caso para defender a tese que esta noção de poder explanatório é, de fato, epistemicamente irrelevante – a não ser, possivelmente, quando combinada com outras virtudes teóricas. Nossa conclusão é que o critério para determinar qual é a *melhor* explicação requer que se leve em conta outras virtudes teóricas além de poder explanatório, como simplicidade e unificação. Neste caso, uma avaliação completa da IBE requer o exame de *se, quando, e como* estas supostas virtudes teóricas são realmente verocondutoras.

Palavras-chave: Filosofia da ciência; escolha de teorias; inferência à melhor explicação

TABLE OF CONTENTS

| | |
|---|----|
| <i>Introduction</i> | 1 |
| PART I. A Study of Reasoning: The Structure of Inference to the Best Explanation | |
| 1. Inductive inferences | 4 |
| 2. The Inference to the Best Explanation | 10 |
| PART II. Reasons Why: Metaphysical Explanation and Explanatory Power | |
| 1. Reasons why | 16 |
| 2. Causal explanation | 21 |
| 3. Metaphysical grounding | 27 |
| 4. Explanatory power | 32 |
| 5. The wonder tissue argument | 36 |
| Appendix. The Inference Against a Non-Explanation (IAN) | 40 |
| Conclusion | 42 |
| Bibliography | 43 |

INTRODUCTION

The Inference to the Best Explanation (henceforth IBE) is an inductive inference, in a sense to be defined. It appeals to the difficult notion of an *explanation* and to an elusive measure of which explanation is *best*. Our aim is to contribute to the evaluation of IBE by evaluating the consequences of an explication of the former notion.

This monograph is organized into two parts. On the first part, our focus is describing IBE. We discuss the nature of induction and outline the challenges of *describing* and *justifying* an inductive mode of inference. Then, we push for a specific formulation of IBE. On the second part, our focus is a component of IBE: the notion of explanation. We begin by explicating a theory of explanation, one that is of special interest to contemporary metaphysicians. It is the theory that *E* explains *F* if and only if *F* holds in virtue of *E*, given appropriate *relata*. We clarify the notion by performing two case studies: *causation* and *grounding* as in-virtue-of relations. We do not argue for that theory. Instead, we argue for a conditional thesis:

If we accept the above theory of explanation, then (i) a certain notion of explanatory power is forced upon us and, consequently, (ii) any inference which appeals only to explanatory power will be inductively weak.

We end with our argument for *ii*, which has been dubbed the *miracle tissue argument*. The upshot of the discussion is that any justifiable IBE must rank explanations by more than just their explanatory power. Common properties appealed to in scientific theory choice are simplicity, unification, and *non ad hocness*. We do not go as far as discussing these in detail.

Before we begin, we provide a broad overview of the epistemological debate about IBE. “Weak explanationism” is the thesis that there can be epistemically justified beliefs obtained through an IBE (LYCAN, 2002). What we may call “strong explanationism” adds that IBE is a basic form of inference, in the sense of not being justifiable solely by appeal to other forms of inference. Strong explanationism considers IBE to be justified nevertheless, and therefore entails weak explanationism. The literature has even expressed what we may call “imperialist explanationism”, whereby IBE provides the justification for all other inductively strong forms of inference, and perhaps even of deductive forms of inference (HARMAN, 1965).

This monograph evaluates *one* critique leveled at *weak* explanationism. Three arguments for skepticism about IBE have become standard in the literature. They are as follows.

S₁) Nancy Cartwright's *simulacrum account of explanation*, motivated by work in the history of science, denies explanatory realism. Explanatory realism asserts that explanation is factive, that is, A cannot explain B unless A obtains. Nancy Cartwright believes the contrary: Given the way scientists employ the concept of explanation, true laws and theories do not really explain, and only overly idealized (and thus literally false) laws and theories explain. Furthermore, what gets explained are events inside a model, not real events. (CARTWRIGHT, 1983.)

S₂) Bas van Fraassen's Dutch book argument. He points to a dissonance between Bayesian probabilities and IBE probabilities. More specifically, he claims that if the standard Bayesianism probability calculus is true, and if degrees of credence match betting behavior, then any pattern of betting behavior motivated by IBE can fall prey to a series of bets that are guaranteed to result in a loss in the long-run – in other words, a Dutch book. *Prima facie*, adopting a probability theory susceptible to Dutch books is thus pragmatically irrational, and this is often held to entail such adoption to be epistemically irrational. The issue, however, is vexed. (VAN FRAASSEN, 1989.)

S₃) The argument from unconceived alternatives, known from Bas van Fraassen's work as the "bad lot argument". Suppose we grant the explanationist contention that the best of all possible explanations is in general true. Still, we can only perform an IBE with the best explanation we have come up with. Yet, quite plausibly, we often fail to come up with the best explanation overall. Unless our best available explanation is somehow compatible with the best explanation overall, it follows from the granted explanationist contention that IBE will fail. The worry here is that this compatibility is rare. (STANFORD, 2006.)

These, however, are not the skeptical worries about explanationism that we wish to address. That worry is that, no matter how much explanatory power may have, this by itself is no indication that it is true. As we have said, we will argue this worry is accurate given a certain notion of explanation. Our result will only be as good as the theory of explanation we rely on.

The above conclusion would not yet refute weak explanationism, for it may still be the case that, when accompanied by theoretical virtues such as simplicity and *non ad hocness*, explanatory power is truth-tropic. We will not go into this latter discussion. However, it pays to introduce the notion of theoretical virtue.

A theoretical virtue is a property which is positively correlated with a theory's truth. Theoretical virtues are grounds for theory choice. This truth-correlation may be context-sensitive, holding only locally rather than globally, in which case we have a local theoretical virtue. Given this definition of what a theoretical virtue is, an increase in theoretical virtuosity is automatically a gain in likelihood (at least in the relevant context). Inferences appealing to theoretically virtuous hypotheses in their premises have an automatic gain in their inductive strength (if the reasoner is aware of being in the relevant context). Note that what seems to be a theoretical virtue might not be so. On this monograph, when we call something a theoretical virtue we will generally intend them to be understood as *alleged* theoretical virtues.

Adequacy to the empirical data is the most obvious form of genuine theoretical virtue. Other properties alleged to be theoretical virtues are not clearly so, and debates on the field of theory choice focus on these other properties. Examples of disputed theoretical virtues are simplicity, elegance, *non ad hocness*, unificatory power, scope, and external coherence.

The chief contention of explanationism is that explanatory power is at least a local theoretical virtue: That, at least in some contexts, a gain in explanatory power translates into a gain in likelihood. We are inclined to accept that, but defending it falls beyond the scope we have chosen for this work. What we will argue for is that explanatory power is not a *global* theoretical virtue. Our argument, in a nutshell, is that explanatory power can be acquired "cheaply", as it were. *Ad hoc* complexifications allow one to explain any data set as powerfully as one wishes. This cheapness makes it epistemically valueless. We have named our defense of this claim the "miracle tissue argument".

Local weak explanationism states that IBE is inductively strong in certain contexts *C*. We can then make IBE globally strong by adding to it the premise that some *C* obtains. One plausible such context *C* is when explanatory power is accompanied by other theoretical virtues, such as simplicity. Thus, theoretical virtuosity could be taken into account in the measure of what makes explanations better than others. The best explanation is then not most explanatorily powerful, but the one that *optimally combines* explanatory power with theoretical virtue. Of course, if a trade-off happens to be needed, a good dose of explanatory power must be retained. Otherwise, the resulting inference would not be explanationist at all.

Having given an overview of the epistemic issues surrounding IBE, we turn to the first part of our monograph. We begin with an exposition of what inductive inferences are, what inductive strength is, and the distinction between global and local inductive strength.

PART I

A STUDY OF REASONING: THE STRUCTURE OF INFERENCE TO THE BEST EXPLANATION

To fully understand Inference to the Best Explanation, we must understand what its inferential structure is, what an explanation amounts to, and would make an explanation the best. Part one is dedicated to examining the first question. Part two moves on to outline a standard answer to the second question, – a certain theory of explanation, – and arguing that it has certain consequences for the third question.

Inductive inferences

The Inference to the Best Explanation is a kind of inductive inference. Inferences are not *basic* sources of positive epistemic status. Their capacity, possibly unique, is *transmitting* positive epistemic status from premises to a conclusion.¹ The better the transmission, the stronger the inference. ‘Positive epistemic status’ is a generic term that encompasses knowledge, warrant, epistemically justified confidence, and similar statuses which the beliefs and conjectures of inquiring reasoners may have. The epistemic status of a proposition is always relative to an agent in an evidential context. Whether the transmissive properties of inferences are likewise relative to evidential contexts is, however, a point of contention.

By ‘inference’ we do not mean inference as a type or a token of a psychological process, but as the content of these processes.² This content is an unordered series of specific propositions P_1, \dots, P_N, C , where C is privileged in that the reasoner takes it to be supported by P_1, \dots, P_N . We may understand a type of inference from this scheme by abstracting from the specificities of each proposition and considering only the types of propositions to which they belong. For example, we can characterize IBE as having premises citing the explanatory power of some theory H of relevant known data, and by having a conclusion asserting the truth of H . Such a generic characterization is what we mean by a “proposition type”.

1 The crisp way of putting this point is due to AUDI (2011: 184).

2 The distinction between inference as process and as content was introduced to us by AUDI (2011: 177).

To specify the inference being performed, we must take into account the type of relation the reasoner holds to obtain between P_1, \dots, P_N and the privileged C . If the reasoner believes there to be a relation of *deductive entailment* (which we will explain later), then he is performing a deductive inference. The very same premises and conclusion, linked by an envisaged relation of *defeasible support*, yield an inductive inference. Of course, in any of these cases, the alleged link may not hold between premises and conclusion. The following inference is inductively strong if R is intended as a relation of enumerative inductive support, but deductively invalid if R is intended as deductive entailment: “That swans s_1 through $s_{1,000,000}$ have been white bears a relation R to the swan $s_{1,000,001}$ being white.”³

Having sketched what inferences are, we may classify them into two broad classes as has been foreshadowed: Deductive and inductive inferences. By ‘inductive’ we mean roughly non-deductive, and we’ll speak of Peircean abduction and Inference to the Best Explanation as inductions in this sense. The contentions of this section will provide a general understanding of how these two and other inductive inferences work.

There are multiple ways to characterize *deductive* inferences. Generically, a deductive inference is one whose premises are intended to guarantee, in some sense, its conclusion. A deductive inference is valid when this intention is satisfied, whereby the conclusion is said to be *entailed* by the premises, and invalid when it is not. The intention is, presumably, the reasoner's intention. There are at least three popular varieties of intended relations of deductive entailment. The generic and agreed-upon requirement for deductive entailment is that the premises in some sense *guarantee* the conclusion. The dispute ranges over (i) whether the scope of this guarantee is metaphysical or purely formal and (ii) whether there is a second requirement of *relevance* between premises and conclusion.

A broad characterization of guarantee can be set out in alethic modal terms: It is not possible for the premises to be all true while the conclusion is false.⁴ What exactly this definition says depends on the scope of possible worlds over which we are defining possibility. If the scope is metaphysically possible worlds, then the following inference is valid: “If water is H_2O then there is no largest prime number.” This is the first popular variety of deductive entailment.⁵ The second variety comes when we take our scope to be logically

3 That is a single proposition, not an inference, of course. We are sacrificing technical rigor for the sake of expository easiness; the example should be clear.

4 One might add a criterion of *non-redundancy*, where the removal of any premise eliminates the guarantee provided by the set of remaining premises. This nicety will be skipped over.

5 Perhaps one could formulate a weaker variety by choosing as a scope the set of nomologically possible worlds. Then we could speak of relations of deductive entailment inside physics even if physical law is

possible worlds, which include worlds in which water is not H₂O (but does not include worlds in which water is not water). Then there must be an (intended) relation of *logical entailment* between premises and conclusion. Now we render deductively valid the inference, “If $P \supset P$, then $\sim (Q \wedge \sim Q)$.” Of course, in reality, what counts as deductively validity will also depend on the logic we adopt. In addition, we may distinguish between syntactic entailment and semantic entailment. Not every logic admits of completeness proofs, so the distinction may be useful. We are not sure if deductive relations inside mathematics would all be syntactical in this sense, given Gödel’s Incompleteness Theorem.

The third and final variety arises when we adopt a *relevance logic*. The last inference we presented relates seemingly unrelated tautologies regarding seemingly unrelated sentences. The premises seem irrelevant to the conclusion, in a sense that is hard to specify. Relevance logicians argue that the relation of deductive entailment can only hold when the premises are (all) relevant to the conclusion. There are many relevance logics, each cashing out the notion of relevance a bit differently, but they share this broad view of deductive entailment. For the sake of completeness, we should mention a fourth kind of logic, a hybrid between the first and the third: One in which the premises must be *metaphysically relevant* to the conclusion, perhaps grounding the conclusion. Theories of deduction employing this view are called, we believe, “hyperintensional logics.” *Guarantee* is necessary but not enough for deductive entailment. Metaphysical relevance, but not traditional logical relevance, is required.

An inference is inductive when its premises are not intended to guarantee their conclusion, but are intended to make the conclusion sufficiently likely. Therefore, induction is not subjected to the criterion of *guarantee*, which is the universally accepted characterizing feature of deduction. Furthermore, deductions are nonampliative and monotonic, whereas inductions are not.⁶ There are, however, two sets of parallel notions that characterize deduction and induction. Let us turn first to these two parallelisms.

The first set comprises three criteria that some wish to impose on valid deductions: non-redundancy, logical relevance, and metaphysical relevance. One can cast these in inductive forms, respectively, in the following way. Every premise in the premise set must be such that:

(i) Were it to be removed from the set, the probability of the conclusion given the set would

contingent. We could also speak of deduction in neuropsychology, even if psycho-physical bridge laws (supposing these to exist) are also contingent.

6 Non-monotonic logics are not standardly called deductive, as their adequacy as models are judged in terms of how well they capture inductive patterns of inference.

decrease; (ii) It is statistically correlated with the conclusion; (iii) It has a causal connection, through finite causal chain in any direction, to the conclusion.

The second set of parallel notions comprises inductive counterparts of deductive entailment and deductive validity. In good inductions, the set of premises should have a *positive inductive relation* to its conclusion. That is, the set must render the conclusion more likely. Standardly, this is formulated in terms of a subjective interpretation of probability, i.e. an interpretation where probabilities are degrees of belief. Since we are inferential realists (i.e., we think inferences aim at truth), we better take the “rationalist” subjective interpretation of probability, according to which probabilities are *rational* degrees of belief (given one’s evidence). We remain neutral on precise accounts of rational probabilities. So good inductions increase the rational confidence levels for their conclusion. The stronger the aforementioned positive inductive relation, the stronger the inference is.

It may be impossible to evaluate the inductive strength of any inference without substantial information about the world. Stathis Psillos (2007) claims, on theoretical and historical grounds, that strong inductive inferences never owe their strength to syntactical relations between premises and conclusion, generalizing the result Nelson Goodman (1955) obtained regarding enumerative induction. Their strength depends on the world behaving in a certain *logically* contingent way (even if its behavior is *metaphysically* necessary). Inductive strength thus depends on the subject matter and circumstantial detail. *Prima facie*, as a result one must evaluate an induction by keeping one’s many other theories of the world in mind. Since one cannot lay out one’s whole web of beliefs in the premises of an induction, the web must be set apart as an implicit “background knowledge” or “evidential context.”

Wrapping up, the inductive counterpart of deductive entailment is a context-sensitive positive inductive relation, understood as an increase in rational confidence, and which we may call *inductive support* or, equivalently, *inductive confirmation*. The inductive counterpart of deductive validity is *inductive strength* but, unlike deductive validity, it comes in degrees. The more an inference’s premises provide inductive support for its conclusion (relative to a background worldview), the inductively stronger the inference is.

Unlike with deductive validity, the relation between inductive strength and transmission of epistemic justification is a bit controversial. Induction is an inference, and inferential transmission of epistemic justification comes from an agent’s reasoning tokenizing this inference (from justified premises). Mentally tokenizing an inference requires *accepting* (or

believing) the conclusion on the basis of the premises (by intending them to have a certain relation of support, as discussed earlier). However, many hold acceptance (belief) to be epistemically justified only when the accepted proposition's rational probability goes over a certain *threshold*. Therefore, certain inductive inferences may be so weak as not to transmit epistemic justification or warrant, despite the fact that their premises *do* increase the conclusion's likelihood. An inductively strong inference can also fail to transmit epistemic justification or warrant if the premises were not very justified to being with.⁷ (We remain neutral on whether inductive inferences are ever capable of transmitting *knowledge*.)

From the above discussion, it follows that induction is ampliative whereas deduction is not. Ampliation, intuitively, is an increase in information. More specifically, an increase in information relative to what was "implicit" in the premises. A deduction's conclusion can regroup information in a form more congenial to our cognition and make us readily notice (i.e. make "explicit") something that was previously "implicit" in the premises. This is not an ampliation in our sense. We will be more precise below.

Another difference which also holds between deduction and induction is that the former is *monotonic*, while the latter is not. Monotonicity is a property of inferences whereby the addition of new information to a premise set P does not exclude any inference one can make from P alone. For example, if we can infer that Q from the premises that P and that P materially implies Q, then no further premise will block this inference. Inductive inferences are non-monotonic: Taking more information into account can alter what is inductively supported by one's premises. For example, one may learn that there are only a million white swans in the world, from which it follows (deductively, in fact) from us having observed a million white swans that the swan $s_{1,000,001}$ will not be white if it exists. Non-monotonic reasoning is also called *defeasible reasoning* for this reason. New evidence can defeat defeasible reasoning, when it blocks inferences justified by the old evidence, or undercut it, when it now renders justified an inference to a contrary conclusion.

Many epistemologists impose a criterion of *total (relevant) evidence* on non-monotonic reasoning. The criterion is that not taking all of one's relevant evidence into account prevents one from obtaining epistemic justification from an inductive inference. The rationale is that

⁷ Note that epistemic justification is not factive, in the sense that one can be epistemically justified to believe a false premise. This is so even if the evidence which grounds one's justification is, itself, factive. Many theorists during this century have taken evidence to be factive, with Timothy Williamson going as far as taking one's evidence to be identical to one's knowledge, a thesis known as $E = K$. (There are some who take epistemic justification to be factive, but this is a highly unorthodox position. See *Justification and the Truth-Connection* by Clayton Littlejohn and its review at the *Notre Dame Philosophical Review*.)

the relevant bit(s) of evidence we already own, but have ignored, may defeat or undercut our inference, and it surely wouldn't be rational to reason while ignoring defeating or undercutting evidence already at hand! (One could claim that ignoring relevant evidence that is *safely* non-defeating or non-undercutting does not threaten epistemically rationality, even if one is not reflexively aware of this safety. One simple example is ignoring the testimony of an unremarkable 1001st witness after unanimous reports from a thousand witnesses.)

Now we can be more precise as to what ampliativity and monotonicity are. If propositions are extensionally equivalent to sets of possible worlds (those in which they are true), then we can model ampliation and non-monotonicity in similar ways. We can model the information contained in a set of premises as the set of possible worlds these premises exclude, i.e. worlds outside their extension.⁸ An inference is ampliative if and only if its set of premises excludes less possible worlds than that set of premises added to its conclusion. The fact that the conclusion eliminates additional possible worlds is an interesting way to see why the conclusion is *not guaranteed* by the premises: The premises are compatible with at least one world excluded by the conclusion, i.e. with a world in which the conclusion is false. The non-monotonicity of induction is illuminated in a similar way: Additional information may show that, probably, one is inside one of those worlds excluded by the conclusion we had obtained earlier. (This is an example of undercutting.)

So much for the general features of inductive inference. We now turn to discuss a specific form of induction: inference to the best explanation. We argue for a specific set of premises, so as to guarantee that IBE has the greatest chance of being inductively strong, all the while having premises knowable or otherwise epistemically accessible by humans.

8 Considerations about scope apply. We may be only considering nomologically, metaphysically, or logically possible worlds. Perhaps we could even speak of epistemically possible worlds, although their precise status *vis-à-vis* metaphysical possibility is very controversial. Due to a lack of technical understanding, we refrain from explaining the use of impossible worlds.

The inference to the best explanation

It would be a mistake to understand induction in general merely as an “Inference to the Likeliest Hypothesis” (ILH). Some inductively strong inferences make no mention of probabilities. They are heuristics, a kind of “interface”, that track probabilistic relations by mentioning other factors.⁹ ILH is trivial, in some sense, and should be our induction of choice whenever possible. Other forms of induction, in stark contrast, are the subject of heated disputes. We engage in them nevertheless because probabilities (absolute and conditional) are not always transparent. In fact, these other inductions are often the only way we could discover certain probabilities. (As a side note, such interfaces are *possible*, we conjecture, due to numerous trial-and-error selections of heuristics during our phylogenetic, cultural, and ontogenetic history. Our inductive practices have evolved.)

It would be equally a mistake to consider our target inference merely as an “Inference to the Likeliest Explanation.” We employ heuristics such as explanatory power, simplicity, *non ad hocness*, unification, and external coherence, – *not* likelihood or relative probability, – in assessing what the “best” explanation is. To be sure, if IBE is to be a good explanation, then the best explanation must tend to be the likeliest explanation. That is, our sense of explanatory goodness, which Peter Lipton (2004) calls “loveliness”, must be fine-tuned to the likely truth.¹⁰ However, at least sometimes we (attempt to) track the likeliest explanation in a roundabout manner: through the interface of the *loveliest* or *best* explanation, the one that intuitively *seems* to bear marks of a true explanation. This has been called “the guiding claim” by Lipton (2004: 124-5). So, anyway, claims weak explanationism.

Describing this interface is a challenge, comprising three tasks: (i) to state the inferential structure of IBE, (ii) to list the criteria we use to assess potential explanatory power, and (iii) to describe how we traded-off explanatory power and other alleged theoretical virtues, such as simplicity and *non ad hocness*, to yield the composite feature dubbed “goodness.” We construct a ranking of better and worse explanations which, hopefully, indirectly tracks true explanations. These are challenges because, as remarked earlier, we often perform inferences

9 The general idea is present in a discussion by Peter Lipton (2004) of the relation between reasoning that explicitly employs the Bayesian probability calculus and reasoning that has explicit explanatory appeals, like IBE. He argues in detail that explanatory considerations are a way to indirectly track Bayesian probabilities, which are too hard directly to assess. The relation between Bayesianism and explanationism is, however, a whole other issue.

10 Peter Lipton links loveliness to our *sense of understanding*. Since we do not know what is the relation between our sense of understanding and genuine understanding, and since we do not know how genuine understanding works and whether it is important, we do not mention *understanding* in our work.

unaware of the heuristics being employed. In this case, it is unclear how we come to judge some explanation as “best”. The best explanation has some optimal trade-off between simplicity (for example) and explanatory power.

One way to obtain conscious access to the principles governing IBE is to observe how it is used by inquisitive humans such as scientists and philosophers, and perhaps even toddlers and police officers. This observation can be either introspective, perceptual, or testimonial. What we can do is twofold. (a) We also detect the conditions under which we infer from such explanatory considerations to conclusions about reality. That is, we discern what considerations (premises) we perform to reach what kinds of conclusions (answering ‘i’). This is what we do here. (b) We also detect patterns in our performing considerations about explanatory goodness over many subjects. With this, we can discern what theoretical virtues we appeal to (answering ‘ii’) and how we negotiate them when they conflict (answering ‘iii’). This task explicates the notions of explanation, explanatory power, and theoretical virtue. Here we explicate only the former two.

Systematizing the rules governing ‘a’ and ‘b’ may be a task no more simple than discovering the rules governing any other complex system, like a computer or a microphysical system, from observing its behavior in varied contexts.¹¹ Stathis Psillos (2007) has emphasized the high degree of context-sensitivity in our inductive inferential practices. Hopefully, these heuristics can be *roughly* expressed in propositional form. They may not be. Sometimes heuristics implemented in connectionist systems, for example, are linguistically inscrutable: They are very intricate mathematical operations *whose basic terms do not denote anything we can recognize as an individual factor*. Let us not concern ourselves with this possibility.

We now move on to characterize a useful version of IBE which is faithful to our inferential practices. Any good inference must have premises which strongly inductively support its conclusion. Premises must also be epistemically accessible to human beings.¹² Inferences transmit epistemic justification from premises to conclusion, so an inference with inaccessible premises is useless. Usually, stronger inductive support translates into weaker epistemic access, and vice versa. So there is a trade-off. The optimal trade-off point may depend on interest and domain. Philosophy may require more easily accessible premises,

11 See LIPTON (2004: 13) for a similar point. Since we’ve already got your attention in this footnote, let us make a marginal point: Our explanatory patterns may evolve over time as we learn and acclimatize to new ways of thinking. The scope of “we” here includes academics from all areas, and it may additionally include all sorts of investigators from outside academia, from toddlers to police officers.

12 By the ‘epistemic accessibility’ of a proposition I mean how difficult it is (for some measure of difficulty) for agents (of a certain class) to be epistemically justified in believing it.

whereas empirical work can handle hard-to-access premises. Our IBE will be modeled on how it would be applied to empirical science.

To begin searching for the canonical version of IBE, let us begin examining an inference-type (hereafter *inference*) so weak we cannot call it IBE.

P1. H is a plausible explanation for E.

P2. E contains our evidence.

C1. Therefore, H.

This has been called *inference to a plausible explanation* by Peter Lipton (2004). However, if explanatory considerations are to be our ground, the availability of a better explanation potentially defeats or undercuts any epistemic justification that a worse explanation could yield. Two applications of the above inference would provide us with conflicting conclusions in almost any scenario. To fix this, we can sharply upgrade **P1** and obtain a genuine IBE. Now we refer to H as *the most plausible* or *best* explanation.

P1. H is the best explanation for E.

P2. E contains our evidence.

C1. Therefore, H.

The first premise is now too demanding. This is related to skeptical worry S_3 : the “bad lot” or “unconceived alternatives” argument. As Bas van Fraassen (1989) and Kyle Stanford (2006) have pointed out, it is *prima facie* highly plausible that, in general, we have not come up with the best of all explanations. Even if we were to be perfect in ranking the explanations we have come up with, the one we judge the best will seldom be the global best. We would thus often fail to be in the position to accept **P1** reliably. Revising **P1** to “H is the best *available* explanation for E” makes it more epistemically accessible.

The chief worry now is that our inference thereby becomes too inductively weak. For the IBE to be inductively strong in the first place, it is necessary that better explanations are usually more likely than worse explanations. So the best explanation is usually more likely than worse explanations. Whenever the best explanation is incompatible with the worse explanations, in most cases, any non-best explanation E will be more likely than not to be *false*. This sorrowful situation will be avoided if and only if one of two things occur. First, some contextual detail renders E more likely than the best explanation. Second, the best explanation is not incompatible with E. This can happen if E is an approximation of the best explanation. For example, classical mechanics is a kind of “prototype” of relativistic mechanics and therefore bears some reasonable degree of truth.

There is some reason to suspect this second scenario is common. It has been pointed out that scientists are often at pains to come up with even a single good theory to match and explain well their data. The reason for this may be quite congenial to explanationists: That it is rare for *false* theories to be both highly explanatory *and* theoretically virtuous. Therefore, if we were to rank theories according to their goodness, – which combines power with virtue, – then the highest ranks would be dominated by true or approximately true theories, which are (of course) significantly compatible with each other. Whether this occurs is an extensive debate in history and philosophy of science we will sidestep. If it does, skeptical worry S_3 is defeated and **P1** yields a good combination of premise-accessibility and inductive strength.

The second premise **P2**, in its turn, is too undemanding. The evidential base E in question may be small, shallow, and contain little variation. Whatever the merits of explanationism, it is widely agreed that the best explanation for *small* (and otherwise unremarkable) evidence sets are not in general true. Therefore, **P2** should be revised to the effect that our evidence base is large, relevantly varied, contains the results of crucial experiments,¹³ and possibly more.¹⁴ It is prudent for us to require E to include our total relevant evidence. Note that, in some circumstances, the set E can be smaller than our *total* relevant evidence. Some evidence can be safely ignored, as discussed above. Below these considerations become **P3** and **P4**.

A minor terminological revision. Some hypotheses *would* explain our evidence *if* they were true. Following LIPTON (2004), we will call those *potential explanations*. The bare term ‘explanation’ will refer to actual explanations. Every actual explanation is a potential explanation, of course. The premises will mention potential explanation to avoid triviality.

P1. H is the best available potential explanation for E.

P2. H is explanatorily powerful.

P3. E is safely close to our total relevant evidence.

P4. Our evidence is numerous, relevantly varied, and includes crucial tests.

C1. Therefore, H.

Notice the addition of a new premise, **P2**. Besides being the relative best, it must also be good in absolute terms. (The best of a bad lot is not good at all.) Furthermore, H’s goodness cannot be *mostly* due to theoretical virtues other than explanatory power. Otherwise, IBE would not be an explanationist inference. Explanatory power must be central to measuring goodness.

13 Crucial tests are normally understood as tests that could distinguish between theory rival theories: An experiment where theories *cross*. Which tests are crucial, then, depends on the epistemic situation.

14 We leave as an open question whether we should understand evidence factively (falsehoods cannot be evidence). The matter is relevant to the epistemic accessibility of our premises. Factive evidence, like knowledge, may not be *reflexively* accessible – that is, discernible based on internally available reasons. The reflexive accessibility of factive evidence and knowledge is compromised under epistemic externalism.

There is a further way to improve IBE's inductive strength without making it too unusable. We can require that we have performed an adequate search for explanations. Adequate search can be understood factively, as a search that makes it objectively likely that we have come up with the actual explanation. This creates a problem of epistemic access: How confident can we rationally be that we have performed an adequate search? Another way is to understand adequate search non-factively. For example, we may say we have searched adequately when competent scientists have put in a lot of effort. As has become a pattern, this eases epistemic access at the same time as it weakens inductive strength, given that our effortful attempts may not be truth-conducive at all.

Finally, the conclusion merits a revision. Suppose our premises are true and H obtains. Still, H might not be the actual explanation of E. Explanation, we believe, requires some connection between *explanandum* and *explanans*. However we understand that connection, the connection may fail to obtain even under favorable conditions. (We will discuss this on Part II.) A good example is *causal preemption*. Consider a toy example. A glass was broken, and the best explanation is that someone threw a rock at it. Well, suppose someone did. However, there is a rather hidden and improbable Rube-Goldberg-like mechanism that, by happenstance, led some other blunt object to break the glass before the rock could. So the rock-throw does not explain the glass-break. (For a discussion, see LIPTON, 2004: 58-9.)

From all these wrinkles and caveats, and suspending judgment about the factivity of some relevant notions, we obtain the following as a canonical version of IBE:

- P1.** H is the best available potential explanation among set S of competing potential explanations for E.
- P2.** H is explanatorily powerful.
- P3.** The set S is the product of an adequate search.
- P4.** E is safely close to our total relevant evidence.
- P5.** Our evidence is numerous, relevantly varied, and includes crucial tests.
- C1.** Therefore, H is true and the actual explanation of E.

In case we already know that H is true, we can that information as a sixth premise and perform the above inference to conclude simply that H is the actual explanation of E.

Recall Stathis Psillos's point that the evaluation of inferences must be topic-specific and context-sensitive (henceforth, just "context-sensitive"). This is something that has been seemingly established about enumerative induction in particular, given Nelson Goodman's grue paradox whereby not all predicates are projectible (generalizable). Given this, we owe an

account of how we could have provided, with a straight face, a general and seemingly context-free form to an inference such as IBE, which is applied in quite “messy” real-life contexts (in Psillos’s words). The answer is that we have formulated IBE with terms that, hopefully, lend themselves to such context-sensitive evaluation. Here are a few examples.

P1 employs the notion of best available potential explanation. We have argued elsewhere that some alleged theoretical virtues humans appeal to in assessing goodness, such as *non ad hocness* and simplicity, *do* have local, context-specific truth-conducibility. Furthermore, quite plausibly humans have learned to employ these theoretically virtuous heuristics precisely because of their truth-conducibility. To do that, then, our goodness judgments would have to track, in a contextually-sensitive way, truth-conducive properties.

P3 mentions *adequate search*. Possibly, our efforts at coming up with hypotheses also reflect our implicit understanding of the world and of relevant circumstantial details, as they do in our enumerative inductive practice. For example, we might have promising guesses about what potential defeaters may be present (PSILLOS, 2007: 443). This leads us to come up with certain hypotheses rather than others in the limited timespan we have available.

Finally, what determines whether our considered evidence E is “safely close” to our total evidence, and whether E is “relevantly varied”, sufficiently numerous, *etc.*, may not be determinable abstractly, but must be determined by the details of the context at hand. Hopefully, we are competent enough at recognizing what the context demands, so that our access to **P4** and **P5** is not too hazy.

Now that we have settled for a canonical version of IBE, after discussing many of its intricacies and possible modifications, we move on to discuss what explanation and explanatory power are. In the end, we offer the miracle tissue argument against the global theoretical virtuosity of explanatory power.

PART II

REASONS WHY: METAPHYSICAL EXPLANATION AND EXPLANATORY POWER

For long, humans have been engaged in explanatory practice. These practices have been the subject of close scrutiny. Whereas the remote origins of methodic *thinking about thinking* are at present unknown, it is known that at least since Aristotle the practice of explaining the world has been the object of philosophical attention. Based on what scientists and other able thinkers have offered as explanations of the world, other competent intellectuals have attempted to explicate (i) what the practice amounts to, (ii) its importance or lack of thereof for human practical interests, and (iii) its epistemological and metaphysical import, which in turn informs the utility of explanation for human *cognitive* interests. Below, we review two broad kinds of theory of explanation: psychologism and non-psychologism. We then explain a distinctively metaphysical conception of explanation within the non-psychologist paradigm and argue that this conception entails a specific conception of explanatory power. We finish by providing the miracle tissue argument to the conclusion that this kind of explanatory power is not a global theoretical virtue.

Reasons why

Throughout the long discussion about this topic, a chief point of dispute is whether the distinctively *scientific* and *philosophical* explanation of things is attributable to reality or if, instead, it is something with mere communicative utility or psychological reality (in a certain sense to be specified). We may, following Alexander Bird (2005), sieve the theorizing about this matter into two broad camps: “subjectivism” and “objectivism” about explanation. For reasons that will be clear, the way we have developed Bird’s categorization recommends that these two terms be substituted by “psychologism” and “anti-psychologism” about explanation. As such, we will often employ the latter pair instead. Let us begin with subjectivism.

To a first approximation, subjectivists regard explanations as mind-dependent. In more detail, subjectivists believe explanations exist only relative to human beings or other agents capable of propositional thought. In even more detail, subjectivists maintain that explanations

are not relations between objects in the external world, – however the external world is understood, – but rather processes that alter one’s way of thinking about the external world. One formulation of subjectivism mentioned by Daniel Nolan (2014: §4) is that “explanation is a matter of what representations have what effects on us.” His formulation rings with our understanding of the term.

Vaguely stated, one early subjectivist family of theories has it that explanations are whatever transform something unfamiliar to us into something familiar, or which otherwise satisfies some sort of cognitive discomfort we have while contemplating a phenomenon. Explanations here may be narratives, diagrams, speech acts, and theories. It should be said that such a reduction in cognitive discomfort should not be obtained through illusion, e.g. through a false theory or a misleading diagram. There can, thus, be some degree of objectivity preserved in the subjective view. However, as Alexander Bird put it, subjectivism holds that explanations do not “constitute part of the way things are.”¹⁵ At best, explanation depends on the way things are in the same way that elegance or color, on standard accounts, do: As a *response-dependent* or *observer-relative* feature of reality. Facts explain each other only insofar as their mention by a competent human communicator can yield the sought-after reduction in cognitive discomfort. There is no mind-independent relation of explanation in *this* sense of “mind-dependence.” This makes clear why we judge “psychologism” an appropriate label for subjectivism.

A similar subjectivist point of view is, roughly, that explanations are that which result in our *understanding* a subject matter. Understanding, here, is not to be understood as an epistemic achievement such as knowledge of explanation. Rather, understanding should be understood in one of two ways. First, as acquiring a *sense* of understanding. A sense of understanding may be nothing but the loss of a cognitive discomfort such as a sense of puzzlement, or perhaps the gain of some new phenomenal aspect. Second, as acquiring pragmatic, discursive, mathematical, or other kinds of *competence* in “dealing” (in some sense) with the subject matter. In this sense one can understand a machine by learning how to operate it; or understand a language by learning how to speak it; or understand a complex theory by seeing how its claims, predictions, and equations fit together.

We have introduced subjectivism as a point of contrast to objectivism, so that its core claims can be understood more vividly. We are not interested here in critiquing these rivals to

15 BIRD (2005: 89).

objectivism. Instead, we want to take up a specific objectivist view of explanation, – which we'll call *inflationary* objectivism, – and explore its consequences. That is, we will argue for conditional theses: If one accepts that inflationary objectivism, then certain things follow about explanatory power, goodness and, finally, Inference to the Best Explanation.

Objectivism (inflationary or not) is not committed to a realist metaphysical theory. Metaphysical realism, broadly construed, states that there are entities whose existence is ontologically independent of us and, furthermore, that some of their properties are also thus independent.¹⁶ Objectivists may well regard the world's constitution as dependent on our minds. The minimal form of objectivism about some x just requires that facts about that x do not depend on some certain aspect of our psychology. In this case, explanatory relations are independent of what propositions have what *cognitive effects* in us. In this precise sense, one may be an objectivist about explanation even if one holds an idealist metaphysics of one's *explananda*. The world may be grounded in one's mind, but the explanatory relations in this world w are independent of any psychological effects which descriptions of w may produce. Let us say, then, that subjectivists are psychologistic and that objectivists are anti-psychologistic about explanation.

There are deflationary varieties of objectivism. Nancy Cartwright's simulacrum account of explanation, reviewed in the introduction (under skeptical worry S_1), is an anti-realist version of objectivism. Explanation exists only within false models. Carl Hempel and Paul Oppenheimer, in 1948, provided an objectivist theory of explanation that achieved a status of orthodoxy for two or three decades in the twentieth century (SALMON, 2006: §1), but which is also deflationary. Although their theory is realist about explanations, explanations are not understood as metaphysical connections (in a sense we'll render more precise).

Theirs is the Deductive-Nomological theory of explanation (henceforth, D-N theory). Without going into too much burdening detail, their account is that the explanation of a particular event is a collection of matters of fact ("initial conditions") together with at least one true universal natural law. Roughly put, together they should be such as to entail the *explanandum*. The explanation is not an inference or argument, but the laws and initial conditions themselves who make themselves susceptible to such an inference. Note that with the same structure one can also specify what is a D-N explanation of laws, rather than particular events. It can even be extended to statistical laws.

¹⁶ For the many ways of formulating metaphysical realism, see *Realism and Anti-realism* (2006), by Stuart Brock and Edwin Mares.

Inflationary objectivism can be introduced by noting three divergences with this standard D-N theory. First, sometimes D-N explanations are *symmetric*. Whenever a certain set of conditions $\{C_1, \dots, C_n\}$ is sufficient for an event E , *given* a set L of laws, one can not only deduce E from $L + \{C_1, \dots, C_n\}$, but also deduce C_k from $E + L + (\{C_1, \dots, C_n\} - C_k)$. Second, sometimes D-N explanations are *reflexive*. One can trivially deductive-nomologically explain a true universal natural law in terms of itself. Third, D-N explanations were given in a milieu where regularist accounts of laws of nature were orthodoxy. On this view, natural laws merely sum up what has so far occurred, and is neither (i) a governing principle that *makes* particular matters of fact be a certain way, nor (ii) an accurate description of some other governing principle, such as an underlying causal power. The laws are ontologically posterior to the subsumed events, which occur ungovernedly. A D-N theory of explanation is compatible with a governing view of laws, but we have called “standard D-N theory” a D-N theory which coupled to a regularist conception of natural law. A standard D-N explanation, therefore, does not require there to be a “metaphysical connection” from *explanans* to *explanandum*.

Some people think that objective explanation is a kind of metaphysical connection. People who think that are *inflationists* about explanation. The reasoning is that the explanation of an x is the *reason why* x obtains.¹⁷ These are not *intentional* or *epistemic* reasons-why, but in some sense *ontological* ones. A seemingly equivalent formulation is that the explanation of something is *that in virtue of which* that something obtains.¹⁸

Standardly, explanations are thought to be irreflexive and asymmetric. No x can be the reason why x itself obtains, and no x can obtain in virtue of itself. Hence the irreflexivity. Furthermore, if x is the reason why y obtains, then y cannot be the reason why x obtains; and if y obtains in virtue of x , then x cannot obtain in virtue of y . Hence the antisymmetry. Coupling irreflexivity to antisymmetry, one obtains asymmetry. The asymmetry of explanation is widely accepted (SCHNIEDER, 2011: fn. 21).

This metaphysical conception of explanation *qua* reason-why, as it is sometimes put (e.g. SKOW, 2016), is objectivist, since it is a standard view that the reason why any p obtains is independent of the psychological effects that p , or its description, would have on agents.

17 When we say that “A is the reason why B”, we are not using ‘reason’ in its intentional sense, as when A is an agent that intended B to obtain and acted on that intention. We are also not using it in its epistemic sense, as when A is evidence for B.

18 We may also say x explains y if and only if y obtains due to x . A fourth manner of speaking is to say that x explains y if and only if “ y because x ” is true, given a certain usage of ‘because’. Benjamin Schnieder (2011) understands certain ‘because’ statements as explanatory. He moves on to develop a logic for ‘because’, but focusing on non-causal explanation.

Causation and metaphysical grounding have been proposed as kinds of reason-why. Before discussing them, let us briefly consider potential counterexamples to the asymmetry of inflationary objective explanation. The proposed asymmetry of in-virtue-of relations may be part of what led theorists to believe their *relata* to be events rather than particulars. At least, if events are understood in Jaegwon Kim’s terms, as the instantiation of a property by a particular at a time. Plausibly, the existence of a particular p at t_0 can be the reason why that same particular p exists at t_1 , given an endurantist view of persistence. That is, once they have begun existing, objects can be the metaphysical basis for their own continued existence. But, by indexing particulars to times, which is partially what events in Kim’s sense do, this reflexive counterexample to the asymmetry of grounding is blocked: the *relata* are distinct events. Below we very briefly sketch two other interesting counter-examples.

The existence of a necessary being has been theorized to be “its own ground” (BLISS & TROGDON, 2016: §6.2). Since grounding is purported to be a metaphysical explanation, this would deny the irreflexivity of explanation. A rebuttal is that such beings would rather seem to be grounded in *metaphysical law*, whatever its nature, or to be simply groundless, obtaining in virtue of nothing. Groundlessness is one way of capturing the phenomenon of fundamentality, and fundamentality one way in which a being can be necessary.

A final counterexample would be causation through closed timelike curves (“time loops” in General Relativity). Some speculative cosmologists theorize their existence. Suppose causation or grounding is *transitive* and suppose that there is a time loop *at the beginning of time*.¹⁹ Then, seemingly, one would have events that obtain in virtue of themselves. If so, then inflationary explanation would be neither antisymmetric nor irreflexive. A rebuttal is that transitivity is implausible if we take the above as substantive metaphysical relations; anyway this is a subject of much controversy (SCHNIEDER, 2011: fn. 22).

These latter two examples trade in very extreme metaphysics. Rather expectedly, our intuition about the reason why things obtain (in specific, that reasons-why are asymmetric) would break down for events that are fundamental in the synchronic and diachronic order of things – that is, respectively, events at the fundamental level at a moment in time and events at the beginning of time. Nevertheless, such relations remain widely accepted as asymmetric, and plausibly they are indeed asymmetric in most cases. We now turn to causal explanation.

¹⁹ Strictly speaking, there would be no beginning of time. One solution to Einstein’s field equations, in a way apparently consistent with the standard Λ CMD model, is a model with time-like closed curves at the beginning of time. It has been discovered by the Princeton astrophysicist John Richard Gott III.

Causal explanation

By ‘theories of causation’, we mean theories of what is understood today as ‘effective causation’ (which might not be what Aristotle himself understood by the Greek counterpart of the term). Many believe that causal relations are explanatory relations in reality. One case is that certain elements in an event’s total causal history are explanatory of that something. Another case is that a complex event or series of events can be explained constitutively in terms of more basic causal relations: a mechanism.

Causation has played a large role in many scientific theories and was an integral part of mechanistic models of concrete (i.e. spatiotemporal) reality. It is central to many theories in the history of philosophy. Nonetheless, during much of the twentieth century, interest in causation declined.²⁰ This is likely due to two factors. First, some philosophers of physics forcefully argued that there is no role for causation in contemporary physics (e.g. RUSSELL, 1912). Second, the intellectual mores among the Anglophone philosophical community was, by and large, an anti-metaphysical brand of empiricism. This led many to accept a form of “deflationism” about causation, if not downright eliminativism. Nowadays interest has been rekindled. Analytic metaphysics has grown sharply and its practitioners have found multiple theoretical and explanatory roles for causation. Furthermore, as far as we know, there has been a leveling of the playing field between philosophers of physics who assert and who deny that causation has a role in quantum and relativistic physics.

Problems about efficient causation can be grouped into three kinds, according to SCHAFFER (2016). These problems were formulated assuming the causal *relata* are events. The formulation of the problems does not seem to depend on a specific theory of events.

First, not every sequence of events bears causal relations. What is required for a causal relation to obtain in such a sequence? Second, causal relatedness seems to have a direction, since causes bear a relation to effects that is not born the other way around. How should we understand this *directionality*? The third arises from a divergence from a common intuition. Some people judge that the obtaining of a causal relation is determined only by the intrinsic features of the causal *relata* and, perhaps, also a governing natural law.²¹ However, the analysis of various cases has led many to accept, as necessary for causation to obtain,

20 Chapter 7 of Bas van Fraassen’s *The Scientific Image* (1980) comments on how discredited that notion was during the time of his writing.

21 See PAUL & HALL (2013: §1.3).

background conditions other than the causal *relata*.²² In these cases, how should one distinguish within a situation between causes and merely background conditions? These are the problems of *connection*, *direction*, and *selection*, in Schaffer's terminology. We are unsure whether it is possible to solve one problem without thereby solving the other two.

There are many models of causation. Some models are deflationary, in a sense. They uphold causation as real, but deny that it is an element of reality that "produces" other elements of reality. Causal relations are not, in these models, reasons why things obtain. They are mere epiphenomena. Other models of causation are inflationary, in a corresponding opposite sense: Causes *do* "produce" their effects and are therefore things *in virtue of which* their effects obtain. The objectivist theory of explanation under discussion does not consider deflationary causation as a kind of explanation. Explanations are held to be metaphysical connections and, no matter how hard it is to unpack the notion of a metaphysical connection, epiphenomena are manifestly not metaphysical connections. Causation must be inflationary. Not every kind of causation called 'inflationary' is indeed inflationary, as we'll see.

Which models are inflationary and which are deflationary? It would be a tall order to review a number of prominent models and assess where they lie in the inflationary-deflationary axis. We will settle with less than that. This is because, fortunately, Jonathan Schaffer (2016) provides a categorization that eases our work. Let us consider theorists who satisfy three conditions. They are realists about causation; they are dealing with the problem of causal connection; and they believe causation can be analyzed (i.e. is not primitive). Such theorists roughly divide themselves into two camps: probabilism and processism.

Probabilism.

According to this position, causation is a kind of "pattern of co-occurrence." Causation has to do with one or another form of strong positive correlation between events. This correlation can be understood statistically, modally, or nomically (with a regularist view of laws). None is fit for causal explanation *as we are understanding it*. An example of the former is Wesley Salmon's model in which C causes E if and only if there is a statistically homogeneous partition of the probability space in which C is statistically relevant to E.²³

22 John L. Mackie's INUS account of causation is an attempt to deal with background conditions: C causes E *iff* C is a necessary but insufficient part of a condition which is, in its turn, sufficient but unnecessary for E.

23 Wesley Salmon (2006: §3.1) provides an introduction to the "statistical relevance" (S-R) model of explanation. For another example, see Patrick Suppes's *A Probabilistic Theory of Causality* (1970). Note that statistics and probability can be understood modally too.

The simplest example of a modal account has it that C causes E if and only if (a) if C were to obtain, E would obtain, and (b) if C were not to obtain, E would not obtain. Two things which satisfy *a* and *b* are said to be *counterfactually dependent*. There are more complicated versions, such as David Lewis's, which attempt to avoid problems such as causal overdetermination and causal preemption.

Sometimes causation is understood as a nomic relation between cause and effect. If the account is to be understood as probabilistic, the notion of lawhood it employs must also be probabilistic. Nomic connections must be held as nothing but certain patterns of co-occurrence (which can, in turn, be understood either statistically or modally). These are known as *regularist* (or Humean) accounts of natural law.

Wesley Salmon was an important theorist of causal explanation. He provided a three-way distinction between “three fundamental views on [causal] explanation” (2006: §4.2) which have “dominated the discussion of scientific explanation from the time of Aristotle to the present” (2006: 62). We were puzzled to notice that his three-way distinction misses entirely any non-probabilistic account of causation. Perhaps this view of history reflects the deep entrenchment that deflationary view of causation apparently had just fifty years ago.

He divides into *ontic*, *epistemic*, and *modal* views of causation. The ontic view has it that explanations of P are worldly facts (or reports of these facts) *conditional on which* P has high(er) objective probability. This is an objectivist and probabilistic account of explanation. Specific ontic views will depend on specific theories of objective probability. Below we will see examples of causation in which the cause reduces the probability of the effect.

The epistemic view is that P is explained by whatever E renders P (more) rationally expectable. That is, the degree of rational credence in “P given E” is high or at least increased. Presumably, rational degrees of credence are always relative to an epistemic context (in which an agent might partake). This account is also objectivist (i.e. anti-psychologistic) and probabilistic, since degrees of rational credence are independent of the cognitive effects of propositions and representations. Yet it is not inflationary. Neither epistemic relations nor patterns of co-occurrence are metaphysical connections.

Finally, the modal view is that P is explained by whatever E necessitates P, in the Carnapian intensional sense that P is true in all possible worlds in which E is true. This is also clearly anti-psychologistic and “probabilistic” (i.e. set out in terms of patterns of co-occurrence). Salmon is not referring to a metaphysically “thick” notion of necessitation.

Processism.

The other broad position suggested by Schaffer sees causation as the occurrence of a certain physical process which results in the effect. The cause will be a selected factor in the physical process. Schaffer (2016: §2) sums up this view by saying “causing is physical producing.” The big challenge here is specifying, non-circularly, what a physical process or a physical production amounts to. Some specifications end up with a “patterns of co-occurrence” view of physical production, in which case processism collapses into probabilism.

Unlike probabilism, processism allows that causes decrease the probability of their effects. Here is an example, due to Wesley Salmon and retold in by Nancy Cartwright (1983: 25). Suppose bits from equally large samples Uranium-238 and Polonium-214 are drawn from in a random manner and then put before a Geiger counter (a kind of radiation detector). Suppose they have a probability of decaying, at any given moment, of 10% and 90%, respectively. When they do decay, the Geiger counter is intuitively *caused* to beep. Since samples are being drawn randomly from equinumerous samples, the total probability of the Geiger counter beeping at any given time is $0.5 \cdot 0.9 + 0.5 \cdot 0.1$, which equals 0.5. However, the probability of there being a beep *conditional* on an Uranium-238 sample having been drawn is merely 0.1, lower than our prior probability. Nevertheless, intuitively the Uranium decay *does* cause the beep. Processist views are usually able to capture this fact by noting that there is a physical process linking the Uranium decay with the beep. On probabilist views, on the other hand, a factor which decreases the probability of an event could never be a cause of that event.

The central theoretical issue for processist views is: What distinguishes genuine physical processes from “pseudo-processes”, such as the spatiotemporal continuous path which shadows traverse on surfaces? (MAYES, n.d.: §4a.) One demarcation criterion is that genuine physical processes are sequences of causally linked events. The shadow’s movement is a pseudo-process because each step in the movement does not cause the next. The account is appealing, but it makes processism fail as an account of causation.

Other three demarcation criteria constitute the views called *interventionism* (or ‘manipulationism’), *transferentism*, and what we may call *strict processism*.²⁴ We will argue that the former two are inadequate as accounts of causal explanation.

²⁴ Strict processist views partake in what is called “process philosophy”, whose metaphysics is built out of processes rather than static 3D particulars or static temporal stages of 4D particulars. We do not know much of this view. Not all processist views partake in process philosophy, in Schaffer’s usage.

Some accounts proposed as interventionist are really *criteria* for the presence of causes. We are interested in the ones proposed as *definitions* of causation. Interventionism, broadly, states: Whatever elements C_1 of a system can be *interfered with* by an “appropriately exogenous cause” C_2 , – not necessarily by an intentional agent, – so as to yield effects which are *in principle* predictable (WOODWARD, 2016). So, if an event E can be reliably indirectly manipulated through a direct manipulation of another event C_1 , at least in a certain context, then C_1 is a cause of E (in that context). Interventionism has become a large, mathematized, multidisciplinary research program. Its models are actively tested for predictive power.

Two difficulties of this view, as Woodward recounts in the aforementioned article, are that (i) some causes may inhabit systems which are in principle not interferable with, and that (ii) the very notion of *interference* is only comprehensible in terms of causation (e.g. as *exogenous cause*), leading to circularity. Even dismissing these problems, interventionism seems to be just another “pattern of co-occurrence” definition of causation. All that is said is that there is a modal relation between events of direct intervention and certain other events.

Transferentist accounts also seem to be mere “pattern of co-occurrence” views. According to Douglas Ehring (1986), transferentist theories hold that causation occurs if and only if (a) two bodies are in spatiotemporal contact, (b) they possess quantitative properties of the same kind, and (c) one body loses a quantity that is then gained by the other. Examples include “velocity, momentum, kinetic energy, [and] heat.” Furthermore, it must be the case that (d) such a change in quantities in each body would not occur except for the presence of the other body. These are just modal relations. Perhaps we have not done justice to the view. There may be other formulations which have a metaphysically thick notion of transference.

Finally, strict processism is part of the overarching program of *process philosophy* (described in SEIBT, 2017). We were unable to find a canonical and precise statement of what processes are, but the idea is that they are primitive “developments” or “activities.” Causation is then seen as a primitive kind of development or activity. This seems like the metaphysical notion of causation that is needed for the objectivist model explanation we are interested in.

The apparent failure of interventionism and transferentism to provide non-circular accounts of causation is understandable. Notions such as “physical process” and “physical production” are very similar to causation. There is no reason to expect the former two to be conceptually or ontologically prior to the latter.

Their failure to escape the “patterns of co-occurrence” paradigm is also understandable. These notions are as opaque as causation is. A similar phenomenon occurs in the analysis of notions such as “making”, “metaphysical connection”, “necessitation”, and “supervenience”. What are intuitively metaphysically thick notions are analyzed merely statistically or modally, leaving many metaphysicians dissatisfied.

Furthermore, as in the analysis of knowledge, there has been a proliferation of “preemption-style” counterexamples to all attempts of analysis. Whenever the relation whereby A causes B is ontologically *posterior* to both A and B, – whereby A and B must satisfy a certain pattern of co-occurrence which *antecedes* their bearing a causal relation, – one can contrive a scenario in which A and B satisfy this pattern *by accident*, without really bearing a causal relation. This has led many to start taking certain metaphysical relations as irreducible to statistics or modality. These relations are thus called *hyperintensional*.

Primitivism and conclusion.

Some realists about causation do not think it can be completely analyzed. That is, there is no successful analysis in terms of necessary and sufficient conditions for causation – unless these conditions mention causation itself or some notion derivative of it. If causal primitivism is correct, then there would seem to be an explanatory relation that is a fundamental aspect of reality. Causes are primitively connected with their effect. Perhaps all explanatory relations, that is, all *in-virtue-of relations* or *reasons-why*, are fundamental. This strikes us as plausible.

The upshot of the discussion, we think, is this. Causal explanation, in the sense of ‘explanation’ we are exploring, requires that causes constitute reasons why their effects occur. *Brute* patterns of co-occurrence seem to render causes explanatorily inert relative to their effects. What is needed for reality’s structure to be explainable are governing laws, fundamental dispositions, strict processes, nonregularist causation, and other varieties of metaphysically thick relations. Their metaphysical thickness cannot be analyzed away, so to speak.

Having made these sadly obscure but suggestive remarks, we leave our superficial discussion of causation aware of its radical provisionality. We now move to the second variety of reason-why we have set out to discuss: *metaphysical grounding*. In this work, we will not touch upon other alleged kinds of reason-why. Some of these are: (i) mathematical reasons-why, (ii) statistical reasons-why, (iii) reasons-why that surface in the study of emergence, dynamic systems, and other complex systems, and (iv) normative reasons-why.

Metaphysical grounding

We have been discussing in-virtue-of relations or, equivalently, metaphysical explanation. One such relation is causation, depending on how it is modeled. Seemingly, there are non-causal forms of metaphysical explanation. What follows is a list of reasons-why that has been proposed in the literature. Some of these have been discussed in LANGE (2016). We will discuss only those of the last kind.

(1) Essential constitution and composition. The essence of electrons explains why electrons have charge. The essences of water and H₂O explain why they bear an identity relation or, at least, a constitution relation. The essence of {*x*} explains why it is essentially mereologically composed by *x*.

(2) Logical-mathematical reasons. Some (but possibly not all) proofs explain why their theorems hold. For example, why there is an infinity of primes, or why geodesic paths appear curved given certain conditions.

(3) Statistical reasons. The law of regression to the mean explains why most sequels to great movies are worse.

(4) An assortment of other reasons-why. Intentions may explain intelligent action. Natural functions may explain normative properties. Natural functions may also explain the veridicality conditions of representations. Boundary conditions may explain the temporal persistence of a system, such as a convection current. Absences may explain why certain events occur, as when the absence of watering explains a plant's death.

(5) Metaphysical grounding. Perhaps subvenient facts ground supervenient facts, and thus explain them. Relations are sometimes grounded in, and thus explained by, the intrinsic properties of its *relata*. Truths are grounded in the world; hence the truthmaking relation is seen as an explanatory relation. A drop in the mean kinetic energy of particles grounds and explains a drop in temperature (SKOW, 2016: 29).

Causation and grounding appear to have something in common: Their *relata* end up with an in-virtue-of relation. Caused events obtain in virtue of their causes. Grounded events obtain in virtue of their grounds. Some have proposed a strong analogy between causation and

grounding. Jonathan Schaffer (2012), for example, claims that causation is a diachronic linkage of distinct existences, whereas grounding is a synchronic linkage of distinct existences.

Others think they are strongly disanalogous, such as Sara Bernstein (2016). If so, then the in-virtue-of relations they create are also of different kinds. Instead of being a natural kind term, the term ‘in-virtue-of relation’ could denote merely a *family* of very similar relations.²⁵ Unfortunately, we have but a metaphorical or imagistic grasp of this relation or relation family. Metaphors such as “linking” and “connecting”, “producing” and “making”, “building” and “transferring”, *etc.* are common, but difficult to cash out. We are unable to explicate them. Partly due to this difficulty and partly due to the proliferation of contrived counterexamples (as in the analyses of knowledge and causation) have led some people to become primitivists about these notions and limit themselves to investigating (i) their structural and logical features, (ii) the strengths and weaknesses of theories employing them, and (iii) whether they are identical and, if not, in what ways they are similar.

Grounding is a theoretical notion designed to capture the judgment that certain things *A* “owe” their very existence to the existence of something else *B*, whereby the *A*’s are grounded in the *B*’s. Let us call this *metaphysical dependence*. Perhaps the *A*’s are token-identical to the *B*’s, or perhaps constituted or composed of the *B*’s, or perhaps they are bear no *ontological overlapping* but are nevertheless related in this special manner. According to Jessica Wilson (2014), what is distinctive about theories of grounding is that all these varieties of metaphysical dependence, called small-g “grounding”, constitute a single notion, which she calls capital-G “Grounding”.

She argues that there is no theoretical role for positing such a comprehensive category: it unifies relations that are too distinct from one another, becoming too broad to do any metaphysical explanatory work (WILSON, 2014: 539-40). We are too unfamiliarized with the discussion to pass any judgment. Below we discuss three cases of small-g grounding and argue that they share certain features. We then note some analogies with causation.

Example one: On truth. Many and perhaps most theorists of truth believe that at least *some* truths are explained by what there is. Very few believe the other way around, – that what is true explains what there is, – although some argue that truth and reality are never related at all. Granting that common view, we have that what exists grounds what is true. Perhaps we wish to take sentences or statements (that is, an utterance of sentence-tokens) as the primary

²⁵ Sometimes crucial terms in metaphysics denote only families of properties or relations. For example, as LOWE (2015) argues, “ontological dependence” denotes a family of relations.

bearers of truth-value. Then what grounds what is true is that certain things exist and that certain sentences have been uttered. If we take the primary bearers of truth to be propositions, and propositions to exist only when mentally tokenized, then what grounds what is true is that certain things exist and that certain thoughts have been thought. And so on.

Example two: On generalizations and their instances. For brevity, we will sidestep issues regarding (a) quantifier interpretation, (b) totality facts, and (c) priority monism's Whole. A (philosopher's) common sense view of quantification is that existential and universal facts are metaphysically explained by what there is. The existence of x grounds the fact that *something* exists. Supposing also the fact that x_1, \dots, x_N are all there is, it being the case that x_1, \dots, x_N are all F's grounds the fact that *everything* is F.

Example three: On wholes and their parts. Priority pluralism is the thesis in hierarchical metaphysics²⁶ that all wholes are grounded in the parts of which they are composed. These wholes may be understood as the (unstructured) *mereological sum* of their parts. Alternatively, they may be a *structured* whole, in which part-part relations help individuate them.

The first of these examples allow us to highlight two features of grounding relations. The first is that they are not amenable to either extensional or intensional analyses. Benjamin Schnieder (2011) points out and that this is so even outside the standard opaque context of a propositional attitude, a contention for which we will argue. To explain why, we will have to explain what an intensional analysis is.

We employ the Fregean analysis of the extension of various terms and the Carnapian model of intension.²⁷ Let w be any given possible world. The extension of a singular term in w , such as a name or indexical, is what the inhabitant of w refers it to: a particular. The extension at w of a predicate is the set of objects to which it applies at w . The extension of a definite description at w is whatever uniquely satisfies it at w , if anything. The extension of a sentence is its truth-value at w , if any. Now, the intension of any term at w is a function from the set of all possible worlds (accessible from w) to the extension of that term at each possible world. Two terms are co-extensional or co-intensional at w when, respectively, they share extension or intension at all possible worlds (accessible from w).

²⁶ *Flat* metaphysics aims to describe what there is and what properties it has. *Hierarchical* (or *ordered*) metaphysics aims to understand which things are prior and which are posterior. Metaphorically, it aims to understand the *structure* of reality, how everything metaphysically hangs together. The enterprise of discovering *reasons why* seems to be hierarchical metaphysics. The terminology is from SCHAFFER (2012: §1.3).

²⁷ The most problematic of Frege's analysis is the common extension attributed to all non-denoting terms, such as the null set. Any chosen object would generate distortions. Consider the resulting truth-value of something like "Mr. Hamlet is a subset of any set." But we will not discuss examples employing empty terms.

The analysis of a term T is extensional or intensional, respectively, if and only if all of the terms in the *analysans* can be substituted for co-extensional or co-intensional terms either *salva veritate* or *salva derivatione*, depending on the type of semantics we prefer (SCHNIEDER, 2011; SHER, 2018). A *salva veritate* substitution preserves the truth-value of any sentence in which T occurs, relative to any world. A *salva derivatione* substitution preserves the inferential relations of any sentence in which T occurs, relative to any world.

A term is hyperintensional in case it cannot be analyzed in intensional terms. (Unanalysable terms have one *analysans*, viz., themselves, and thus are eligible to being intensional.) A proof that some kind of grounding is hyperintensional requires a case study. Let us examine the case of truthmaking. Any case study of truthmaking would require a specific ontology, a precise notation and a semantics of truth; our lack of expertise will force us to be accordingly imprecise. Suppose we accept two theses: That propositions are the primary truth-bearers and that they necessarily exist. Let “ $\ulcorner P \urcorner$ ” denote the proposition *that P*.

We may say truly: “ x ’s being colored is grounded in x ’s being red”. Since we intend to discuss truth, we may alternatively say (perhaps also truly): “ $\ulcorner x$ is colored \urcorner ’s being true is grounded in $\ulcorner x$ is red \urcorner ’s being true”. Now, suppose we can say truly:

(G) “ $\ulcorner P \urcorner$ ’s being true is grounded in the world being such that P ”.²⁸

The *relata* in the predication (G) are: A state of affairs (a proposition’s having the property of being true) and another state of affairs (the part of the world wherein P obtains). Sentences expressing these states of affairs are predications and therefore complete sentences. Their extension at any w is then their truth-value at w . These two sentences have the same truth-value at every w (and, *a fortiori*, at every w accessible from the actual world). Therefore, they share extension at every w . As a result, they are co-intensional.

If the grounding relation were an intensional predicate, then we could perform the following substitution *salva veritate* on (G): “the world being such that P is grounded in $\ulcorner P \urcorner$ ’s being true”. We may safely take this to be a falsehood (a necessary one, even). Therefore, the grounding relation is hyperintensional. More examples of hyperintensionality can be found by considering relations between necessities. Mathematical necessities may bear grounding relations between themselves but not bear such relations towards metaphysical necessities (such as water’s being H_2O whenever it exists).

28 The roundabout locution “the world being such that” seems required to preserve grammaticality.

The second feature worth highlighting is that grounding, like causation, seems to admit of both overdetermination and partiality. ‘There is something’³s being true seems to be (completely) grounded on each and every existent. We are not aware of skepticism about such overdetermination. Causal overdetermination, by contrast, is highly contentious. For example, in the transference theory of causation, a single cause imparts some quantity of a certain property (e.g. energy) into some other object. Two causes would impart more of that property, which is a different effect. Overdetermination, however, requires that one of the causes be sufficient for the entire effect. So causal overdetermination is impossible on that view. Other views, seemingly, discover other difficulties with causal determination.

Now onto incompleteness. Like partial (i.e. incomplete) causes seem possible, partial grounds also seem possible.²⁹ The truth of ‘P’ partially grounds the truth of ‘P ∧ Q’, supposing that ‘Q’ is also true. (Seemingly, partially or not, ‘P’ could not ground anything which did not obtain.) Each part of a whole partially grounds the whole itself, given priority pluralism. Finally, whenever a universal generalization is true, it is partially grounded by the elements in its domain (or the truth of its substitutional instances).

This brings us to the question of whether complete grounds necessitate what they ground. Partial grounds do not modally necessitate their groundee unless they necessarily co-occur with other partial or complete grounds for that groundee. Likewise, partial causes, such as “tendencies” (as in the propensity theory of probability) or “dispositions” (as in dispositional essentialism), do not necessitate their effects. But that *complete* grounds necessitate their groundees seems to be a common view. Yet, Ricki Bliss and Kelly Trogon (2014: §3) mention a few articles from the past decade which dispute such necessitation. Grounding and modality would then, indeed, bear no straightforward relationship.

Despite these two analogies between causation and grounding, viz., the possibilities of overdetermination and partiality, and despite their being both sources of apparently the same kind of in-virtue-of relation, there is a potentially strong disanalogy. Whereas many take causation to obtain only between events, many accept that grounding can occur among *abstracta*, states of affairs, events, and perhaps also entities from other types. Furthermore, it seems that grounding can occur between items of different ontological types, such as when worldly facts ground the truth of abstract propositions (BLISS & TROGDON, 2014: §3). Perhaps, though, we have been unduly restrictive regarding causation.

²⁹ For more on these, see BLISS & TROGDON (2014: §5).

Explanatory power

Forward problems are about discovering the effects of known conditions. The inference to the best explanation attempts to solve a *reverse problem*: discovering the reason why a known effect occurred. Reverse problems are harder because many different initial conditions can lead to the same end result. To carry out the task of discovering true explanations, IBE employs a notion of a “best potential explanation”. In this section, we better explicate what better explanations are. Before beginning, let us organize our work relative to the questions in which a student of explanation can be interested in. Given a data set E , we ask:

- i. *Definition*. When is some H (potentially) explanatory of E ?
- ii. *Comparison*. When does some H_1 (potentially) explain E more than some H_2 ?
- iii. *Confirmation*. When is some H 's (potentially) explaining E confirmatory of H ?

At the beginning of Part II, we defined a notion of explanation and thus answered question *i*. To get clear on what our answer meant, we saw the applied cases of grounding and causation. On this section, we will argue that the above notion of explanation entails a certain notion of *explanatory power*, forcing on us an answer to question *ii*. Explanatory power is how much something explains.

Question *iii* matters the most for students of IBE. The other two are stepping stones. We lack space to go deeply into it, but tackle it briefly twice. First, we examine whether one aspect of explanatory power (namely, *scope*) is confirmatory. Second, we offer the so-called miracle tissue argument, for the conclusion that explanatory power is not confirmatory *by itself*. Question *iii* is probed in more detail in another work.

We move on to answering the second question. Explanatory power is an essential part of what it is to be the *best* potential explanation. We will later argue it is not the only part, and that other theoretical virtues must be considered. Now, we argue these two measures of explanatory power are forced on us by the conception of explanation as a reason-why: *completeness* and *scope*. One is qualitative, the other quantitative.

Let us begin with completeness, since we have just done preliminary work for this on the last page of the previous section. There are two ways to define the complete explanation of E :

- (D1) The set of all reasons why E .
- (D2) Any set of reasons why E which necessitates it.

There are two ways for these definitions to diverge. First, in case *E* is necessitated by many sets of reasons, i.e., *E* is overdetermined. D1 says only the set of these sets is the complete explanation of *E*. D2 says any of these sets are already complete explanations of *E*. Second, in case *E* is not necessitated by anything, i.e., *E* is underdetermined). D1 says the set of all the insufficient reasons why *E* occurs is the complete explanation of *E*. D2 says it has no complete explanation. This dispute seems a matter of definition. We arbitrarily opt for D2. Should there be chance events in reality, then we say there is no complete explanation of that event. One who speaks in the way of D1 would say that the complete explanation of the chance event would be “faulty” or “permissive”: It could occur without that event occurring.

So, the more a hypothesis *H* approximates a complete explanation of *E*, – i.e., the more elements of *E* it explains and the closer they are to being necessitated, – the more it explains *E*, and thus the more explanatorily powerful it is with respect to *E*. Hence, *completeness* is a measure of explanatory power. Underdetermined events do not have complete explanations.

One surprising result is that hypotheses which appeal to unlikely coincidences are not *weaker* explanations for that – although they might well be *worse* explanations insofar as they are unlikely.³⁰ At any rate, they are not *less complete* explanations. A cosmic confluence of factors may well be the complete reason why something occurs, like the statistically delicate contraptions of a Rube Goldberg machine can completely cause its end result.

We finish the discussion of explanatory completeness by stating two unsolved problems. First: *Ceteris paribus* clauses state that something is the case so long as certain conditions are respected. These clauses do not explicitly specify what conditions these are. Thus, explanations involving *ceteris paribus* clauses fail to explicitly state the relevant reasons-why. Should we consider hypotheses with such clauses to be incomplete? In a certain sense, the *ceteris paribus* clause elliptically refers to each and every reason-why, since, presumably, the truth-conditions of sentences containing *ceteris paribus* clauses involve the omitted conditions. Perhaps, then, they should be regarded as *elliptical* complete explanations.

Second: Can absences be reasons why something obtains? There are two sets of examples which *prima facie* have absences as reasons-why. The first contains examples from everyday life: The absence of a cookie in the jar is a reason why some kid is disappointed; the absence of watering is a reason why some plant died; an ominous shadow (the absence of light) is a

³⁰ Keep in mind our distinction between sheer explanatory power and *loveliness* (or explanatory goodness). The latter is a measure constructed to serve the epistemic purposes of IBE: discovering true explanations. Strong but unlikely explanations are thus *bad* explanations.

reason why someone became worried. The second set of examples intrudes itself on cases where *presences* (rather than absences) are reasons-why. Suppose we have some *C* standing as a reason why *E* obtains. It seems that, in giving the complete explanation of *E*, we must mention the absence of any intervening factor that could have severed the connection between *C* and *E*, for some appropriate modal scope. For example, the absence of finks, in the theory of dispositions, is a reason why a stimulus led to a manifestation. The absence of defeaters, in the theory of justification, is a reason why someone is justified in some belief. The absence of preemptive causes, in the theory of causation, is a reason why something caused its effect.

Perhaps these can be restated only in terms of what is there, rather than what is not there. If they cannot, then absences would seem capable of being reasons-why. The result is strange because of our “thick” metaphysical view of reasons-why. How can that which is not there ever *make* something happen or be the case, like causation and grounding do? The question is as vague as questions in metaphysics get, and we do not know how to begin answering.

The second criterion is *scope*. It is a quantitative measure. Completeness measures how well *each* empirical fact is explained. Scope measures how *many* empirical facts are explained. There are some subtleties, however. We must not ignore the variety of *explananda*. A hypothesis which explains very similar things does not have explanatory power. It explains numerous times the same type of thing, but tokenized at different times. True quantity of *explanandum* requires variety, either in the conditions of measurement or in the measured objects. Note we will not get into the metaphysical mare’s nest of defining the criteria for two data being “qualitatively different” (or belonging to different “classes of facts”). We will take for granted an intuitive understanding of this notion.

Having defined completeness and scope, we have finished our answer to question *ii*. We now briefly comment on whether explanatory power is confirmatory. There is reason to think that it is so. Not, however, unrestrictedly – the miracle tissue argument would prohibit it. When choosing among our *plausible* theories, their scope can be epistemically revealing.³¹

Unfortunately, not every kind of scope will be relevant. For instance, holding other things constant, it is epistemically inert to introduce variations in daytime when doing experiments in chemistry. Paul Horwich (1982: ch. 6) has an ingenious account. Consider that, if there were a plausible theory according to which variations in time can indeed affect an

³¹ Plausibility may be partially a function of theoretical virtuosity. For example, coherence with our background theories (which is not just logical consistency with them). Plausibility is also determined by how it fits with our observations, of course.

experiment's outcome, – say, because physical constants fluctuate over time, – then suddenly it becomes epistemically relevant to have time-variety in the data. Among plausible theories, a theory can be confirmed by being consistent with, predicting, or even explaining the data which is inconsistent with, counter-predicted by, or unexplainable by many rival theories.

Horwich expands this idea into a general theory: The epistemic relevance for a theory T of variation in data relative to a certain factor F depends on what plausible rival theories there are. As foreshadowed, a few conditions should be met. First, some predictions of most of these theories should be to some degree a function of F. Second, if such variation is to be capable of improving T's epistemic status relative to its rivals, then these theories must, for the most part, disagree with T about the outcomes of such variation. A wider variety of classes of fact is epistemically relevant insofar as it is capable of refuting most of the plausible theories about the domain in question. Horwich provides a simple Bayesian proof of the epistemic relevance of this kind of scope, which we will not rehearse. If he is correct, then *this* kind of explanatory scope is confirmatory insofar as it is accompanied with the parallel kind of predictive scope which Horwich discusses.

We conclude by asking: Do the alleged theoretical virtues of *simplicity* and *unification* increase explanatory power? We owe an explanation of why we did not include them in this section. We believe that they indeed contribute to goodness and are, as some say, “explanatory virtues” in this sense. However, they are so only because they seem to be theoretical virtues. They seem to be properties typical of true explanations and untypical of false ones, but that's it. As we have seen, explanatory power increases as a theory postulates more complete explanations (i.e. stronger reasons-why) for more classes of facts (i.e. more reasons-why). There seems to be no other tenable conception of explanatory power.

Simplicity, however defined, is orthogonal to completeness and scope. Unification, on the other hand, is just explaining a many classes of facts in the same way, providing for them the same reason-why *R*. Unification does not increase scope: although it *requires* a wide scope, it does not *contribute* to it. Neither does it increase explanatory completeness: It fits observations into overarching patterns but, since patterns in the sequence of events are posterior to the events themselves, no more complete reasons-why for these events are presented.

Perhaps *depth* of explanation increases completeness: Would some *E* would be explained more completely in case its explanation *R* was itself explained by another reason-why *R**? We will not pursue this, but the matter of explaining explainers will surface tangentially below.

The wonder tissue argument

The more an explanation is theoretically virtuous, the more likely it is – other things held constant. So it makes sense to take into account theoretical virtuosity in assessing the best explanation. It makes sense because it increases the inductive strength of IBE. This is not surprising. What is perhaps surprising is that it also makes sense to sacrifice considerable explanatory power for the sake of theoretical virtuosity. So long as enough explanatory power is preserved, the result of the trade-off can still bear the title of the best *explanation*. This claim raises two questions.

Question one: Epistemically, why is the trade-off worth it?

Question two: Epistemically, why not trade-off *all* explanatory power?

The answer to the first is: Explanatory power is easy to acquire by theft. By ignoring the demands of (alleged) theoretical virtues such as simplicity and *non ad hocness*, one can trivially formulate a theory which *completely* explains *every single datum*, thus scoring maximally in the two criteria for explanatory power: completeness and scope. We will argue for this claim by way of a case study, the “wonder tissue”, which embodies a recipe for creating false yet maximally powerful explanations.³²

An answer to the second we would wish to explore in further work is: Explanatory power is very hard to acquire by honest toil. Should an explanation be simple, *non ad hoc*, coherent, and so forth, perhaps will be at pains to provide solid reasons-why for a large number of data. Consider how difficult scientists find it to discover a *single* virtuous theory which explains (or even predicts, for that matter) their data. So, – one would somehow conclude, – when a virtuous theory is indeed explanatory, this is indicative that it is true.

The wonder tissue argument begins with a hypothesis that explains a data set *E*. The hypothesis is that each of *E*'s constituent data was *necessitated* and *grounded* by a black box entity, who's got an *ad hoc* assortment of causal powers or grounding capacities – and exerts them in some unexplained (but possibly explainable) way. Perhaps it could be something of the form, “For every fact, known or unknown, God did it on a whim”, though

³² “Wonder tissue” is a term coined by Daniel Dennett to denote posited entities with strong metaphysical powers, such as “thinking rationally” (e.g. a black-box faculty of Reason) or “creating the laws of nature” (e.g. a mysterious divinity), but whose *internal workings* are completely unspecified. As a result, the strong metaphysical powers of these entities remain completely *unexplained*. They are strong explainers and, all the while, they are strongly unexplained. Something is fishy about wonder tissues. See chapter 22 of his *Intuition Pumps and Other Tools for Thinking*. New York, NY: W. W. Norton & Company, 2013.

the black box needn't be a divinity. The entity's power is such that no other condition must obtain in order for its powers to take effect; it is the sole source of what it causes or grounds.

Not having detailed the entity's capacities and psychology, we have cheaply posited an unexplained explainer – a wonder tissue. The wonder tissue certainly does not exist and quite plausibly is a metaphysical impossibility. Nevertheless, the wonder tissue hypothesis remains very explanatorily powerful, for reasons explained below. If so, then any version of IBE which ranked potential explanations only according to their explanatory power would always lead us to infer the existence of such a non-existent and perhaps *impossible* wonder tissue. Since IBE is used as a guide both to metaphysical possibility and to actual truth, some revision is forthcoming. One could employ another theory of explanation whose consequent notion of explanatory power cannot be rigged in this way. But, having assumed an inflationary notion of explanation in order for IBE to be reliable we *must* take into account something other than brute explanatory power. That is our thesis.

Now here is why the wonder tissue is greatly explanatory. The black box's actions provide an explanation for each *explanandum* in a data set E , and an explanation which moreover necessitates what is explained. For example, the wonder tissue could have single-handedly and continuously grounded or caused physical objects to act, as a matter of necessity, in accordance with gravitational law. As such, the wonder tissue can be set to *completely* explain anything. It can also be set to explain as many kinds of things as we want, observable or not, making it also an explanation with the maximum possible *scope*. Therefore, it is perfect in terms of explanatory power. We can produce multiple (incompatible) theories of this sort, varying the details of the wonder tissue, and thus proliferate very powerful false explanations. This defeats the truth-conducibility of pure explanatory power.

There is a hidden stumbling block which prevents the maximization of the explanatory scope of our wonder tissue. Let E be any data set, and e_1, \dots, e_N its elements. It is possible that E contains *internal explanatory information*, i.e. observations of the form: " e_x is the reason why e_y ." Some empiricists deny that facts of this form are ever part of our empirical data; they are always part of a theory about the data. If so, so much the better. Otherwise, a black box could not be accepted as the sole reason why some such e_y obtains. The black box hypothesis would contradict E itself. To avoid this, we can shift our hypothesis so that the black box in question has grounded only the observations for which E has no internal explanatory information. This limitation of scope applies to any theory, however, so that the

wonder tissue hypothesis remains towering above other hypotheses in its explanatory power.

Explanatory power, then, is not truth-tropic by itself, i.e., in any context. That is, it is not globally truth-tropic. This spells trouble for IBE, which takes explanatory power as the central measure of explanatory goodness. Otherwise, there would not be much of an *explanation* to the “best explanation”; this is why IBE cannot trade-off explanatory power away indefinitely. This answers ‘question two’ above.

In order for IBE to be reliable, then, we must find some local context *conditioned on which* explanatory power becomes truth-tropic, and apply IBE only under such contexts. For instance, restrictions upon what kinds of explanatory hypothesis we are considering. Perhaps simple, *non ad hoc*, unified, *etc.* hypotheses are a context in which explanatory power indeed becomes strongly positively correlated with truth.

Consider how the miracle tissue hypothesis lacks these alleged theoretical virtues. It includes the postulation of *ad hoc* causal powers and grounding capacities. It completely lacks predictive power. It explains complex phenomena *via* a black box, an unexplained explainer with wonderful capabilities, like a *virtus dormitiva*, which we should receive with skepticism. It is inconsistent with our background theories, as they provide various explanations to our data, whereas the wonder tissue is posited as the sole explanation of everything. Things get worse in case our wonder tissue is some sort of divinity – disembodied minds with irreducible semantic and rational powers are nothing like anything else in our scientific inventory. Notwithstanding, the wonder tissue remains explanatorily powerful. A theoretical vice could only harm explanatory power by diminishing either scope or completeness. *Prima facie*, the above theoretical vices seem not to diminish these two.

In case it still feels difficult to accept the explanatory power of the wonder tissue, consider the following scenario: If one were to suggest this hypothesis in an observational situation, it would be appropriate to say: “Well, that *would* explain it – *but*, it is incredible and *ad hoc*.” The emphasis on the subjunctive mark emphasizes the hypothetical (i.e. unlikely, outlandish) nature of the potential explanation, but we do not consider that such a nature diminishes the theory’s capacity to explain.

Let the *black box equivalence thesis* be that, other things held constant, *unexplained* explainers can explain as well as explained explainers (that is, they have identical explanatory power). One may even accept that *unexplainable* explainers can explain as well as explainable explainers. Now let the *virtue-vice equivalence thesis* be that, other things held constant,

theoretically vicious explanations have as much explanatory *power* as virtuous counterparts. These modalities of explanationism can come in global and local varieties too.

Our discussion of explanatory power has aimed to establish both theses, given the inflationary theory of explanation. This allowed the wonder tissue argument to establish the great numerosity of explanations who are powerful, vicious, black boxy, and *false*. This demonstrates that explanatory power is not globally truth-tropic, since wonder tissues do not exist.³³ To rescue IBE, we must construct our measure of explanatory goodness that *decreases* as theoretical viciousness and black boxness increases, so that the best explanations, while explanatorily powerful, are neither vicious nor black boxes. Whether any such a measure could be itself truth-topical is topic for subsequent investigation.

33 Although, on the PBS documentary *A Glorious Accident* (1993), Freeman Dyson has claimed, *contra* Daniel Dennett, that physicists posit wonder tissues all of the time to do some required causal work. We'd reply that the posited fields and particles are not *as* wonderful as the complex capacities our above hypotheses posit. They're instead quite modest capacities and patterns of behavior, something one could reasonably accept as fundamental.

APPENDIX:
THE INFERENCE AGAINST A NON-EXPLANATION (IAN)

Theories can fail to explain some data. This is inconsequential insofar as the data in question are outside the theory's domain. Theories can, however, fail to explain data in their domain. Insofar as we expect these data to be explainable, the theory in question is thus likely to be *incomplete*. It may nevertheless be true as far as it goes. Trouble comes in two scenarios. First, when a theory renders something explainable only by a cosmic coincidence, when we have reason to think it has a non-accidental explanation. Second, when it makes it likely that something we expect to have an explanation does not, in fact, have an explanation.

We might have called theories of those two kinds *anti-explanations*, but we call them *non-explanations* for the sake of a good acronym to name our inference. If we infer from a theory's being a non-explanation that it is false, we have performed an Inference Against a Non-explanation (IAN). Below we justify this inference-type and consider two applications.

Certain things can be acceptably unexplainable. Baffling as it may be, explanations have to end somewhere. Basic causal powers, metaphysical laws, and logico-mathematical rules are likely to hold for *no reason at all*. Other nonfundamental facts but still hold for no *perspicuous* reason, in the sense of being the products of a *chance* confluence of causes and other reasons-why. A theory which explains them would merely detail the step-wise chain of coincidences which produced them, but does offer any *counterfactually-robust reason* for their occurrence (unless it is supposed that the modal space over which counterfactuals range is also inhabited by worlds with similar chance events).

A third class of events is those neither fundamental (and are thus somehow explainable) nor accidental. We are of the opinion that *complex regularities* pertain to this class. While the Humean metaphysician accepts the nonfundamentality of complex regularities, he believes they are explainable only in terms of chance. We hold this combination of *chance* and *order* to be extremely unlikely. (Even order arising in *chaotic systems* in natural phenomena is not due to chance, but arises due to causal interaction among its parts.) So no adequate theory could make it unlikely or impossible for complex regularities to have a counterfactually-robust and ultimately *non-chancy* reason for their occurrence. If we accept that, there are consequences for our approach to matters in metaphysics. Below we offer two examples.

First, high-level facts about how organisms generate consciousness. The natural laws in question would be too complicated to be simply *primitive*.³⁴ Plus, it couldn't occur merely by chance. The reason is that there is a systematic *match-up* between organismic behavior and phenomenal character – i.e., between phenomenal reports and phenomenality itself. Here are illustrative examples. Phenomenal suffering is correlated with stimulus avoidance: A match between physical and phenomenal aversiveness. The experience of spatially arranged items matches our report of a spatially arranged environment. Finally, the visual experience of *contrasting colors* is correlated with the physical capacity to discriminate colored objects. For example, colorblind people do not experience colored objects as *sharply distinct*, and this is matched by their inability to *discriminate* their physical correlates. Given this surely non-accidental match-up, there must be *mechanisms* or *governing laws* sensitive to phenomenal character and which produce physical effects appropriate to that character. Theories which imply there being no such reasons are rightly discarded by IAN.³⁵

The second example, which has already been foreshadowed, is that certain metaphysical views, such as metaphysical Humeanism, leave the possible sequences of events perfectly unconstrained. Individual events may be constrained by their individual natures, but there is no constraint relative to which distinct events are co-present or temporally sequential. Such a view would make it exceedingly unlikely for reality to exhibit *any* widespread, enduring order – yet, we know that it has exhibited such an order since its very beginning. This is a breathtakingly unlikely event for the Humean. Since the Humean cannot explain non-accidentally this ordering, IAN discards it (and, seemingly, quite justifiedly).

A final example, which we will not expand upon, involves the difficulties that scientific antirealism has in explaining why certain scientific theories are successful while others aren't. Most antirealist attempts at this are at best sketchy (KUKLA, 1996). Since it is to be expected that scientific success is non-accidentally explainable, antirealism is at odds with IAN.

IAN is an inference more modest than IBE. Consider that IAN is weak enough so as not to even reject miracle tissue explanations. While IBE assumes that the world is stratified by powerful and *virtuous* explanations, IAN merely assumes that most of its patterns have *some* non-accidental explanations, rather than being primitive or only accidentally explained.

34 This is not to say physicalism is true. Perhaps these laws linking phenomenal consciousness to access consciousness in complex psychologies may derive from much simpler *primitive* psycho-physical laws linking physical properties (such as information) to phenomenal properties. Such a primitive nonphysical law would defeat physicalism on any of its versions.

35 Explaining this match-up is part of the so-called *hard* problem of consciousness. The problem is hard because theorists have (arguably) not come up with a single *potential* explanation of the match-up short of *denying* it.

CONCLUSION

We often perform inferences without awareness of their governing principles and niceties. The principles and details relevant to deductive inferences have been exhaustively studied, but the study of non-inductive inferences has not fared so well. Proponents of any non-deductive mode of inference face a hard task of *describing* their inference rules, which can rival in difficulty the effort of *justifying* them. Explanationists face a descriptive problem of this kind. Their lack of success so far has led to serious charges of obscurity (LIPTON, 2004).

We have not come to the point of assessing the inductive strength of IBE, tackling the problem of justification. However, we have covered much of the ground leading up to such a valuation, by tackling the problem of description. Different notions of explanation output IBEs with differently inductive strengths. Hopefully, we have also alleviated to some degree charges of obscurity against IBE.

After outlining what theories of explanation aim to do, we have chosen, explicated, and applied a notion of explanation which is of special interest to metaphysicians: in-virtue-of relations or, what is the same thing, reason-why relations. We then argued that this theory of explanation implies a considerably narrow notion of explanatory power, which is seemingly independent of theoretical virtues such as simplicity, *non ad hocness*, and unificatory power. This allowed us to argue that explanatory power can be obtained quite easily by manifestly false theories. As a result, explanatory power *per se* is not truth-tropic. Since the ranking of best explanations used in IBE necessarily puts great weight into explanatory power, this creates the worry that IBE is a terrible inference.

We conclude that a reliable version of IBE requires its premises to state that a certain *local context*, in which explanatory power is indeed truth-tropic, has obtained. One possible such local context is the presence of other (alleged) theoretical virtues in our explanatory hypotheses. Somewhat plausibly, it is a great achievement (and a sign of truth) if a simple, *non ad hoc*, and externally coherent theory achieves great explanatory power. A full assessment of IBE, therefore, requires analyzing how explanatory power fares when conditioned upon the presence of these theoretical virtues.

BIBLIOGRAPHY

- BARNES, Eric C. “Prediction versus Accommodation.” In: ZALTA, Edward N. (ed.). *The Stanford Encyclopedia of Philosophy*, 2018.
- BEEBEE, James R. “The Abductivist Reply to Skepticism.” *Philosophy and Phenomenological Research*, v. 79, n. 3, pp. 605-36, 2009.
- BERNECKER, Sven; PRITCHARD, Duncan (eds.). *The Routledge Companion to Epistemology*. New York, NY: Routledge, 2011.
- BERNSTEIN, Sara. “Grounding is not Causation.” *Philosophical Perspectives, Special Issue: Metaphysics*, v. 30, n. 1, 2016.
- BIRD, Alexander. “Explanation and Metaphysics.” *Synthese*, v. 143, pp. 89-107, 2005.
- BLISS, Ricki; TROGDON, Kelly. “Metaphysical Grounding.” In: ZALTA, Edward N. (ed.). *The Stanford Encyclopedia of Philosophy*, 2014.
- BRAITHWAITE, Richard B. *Scientific Explanation*. Cambridge, UK: Cambridge University Press, 1953.
- BROMBERGER, Sylvain. “Why-Questions.” In: COLODNY, Robert G. (ed.). *Mind and Cosmos*. Pittsburgh, PA: University of Pittsburgh Press, pp. 86-111, 1966.
- CARTWRIGHT, Nancy. *How the Laws of Physics Lie*. Oxford, UK: Clarendon Press, 1983.
- CHAKRAVARTTY, Anjan. “Scientific Realism.” In: ZALTA, Edward N. (ed.). *The Stanford Encyclopedia of Philosophy*, 2017.
- CHARTER, Nick; VITÁNYI, Paul. “Simplicity: A Unifying Principle in Cognitive Science?” *Trends in Cognitive Sciences*, v. 7, n. 1, pp. 19-22, 2003.
- CLARK, Andy. *Being There: Putting Brain, Body, and the World Together Again*. Cambridge, MA: MIT Press, 1997.
- DAWID, Richard. “The Significance of Non-Empirical Confirmation in Fundamental Physics.” Pre-print, 24 pp., 2017. Available at arXiv.org: <https://arxiv.org/abs/1702.01133>.
- DAWID, Richard; HARTMANN, Stephan. “The No Miracles Argument Without the Base Rate Fallacy.” *Synthese*, v. 195, n. 9, pp. 4063–4079, 2018.
- EHRING, Douglas. “The Transference Theory of Causation.” *Synthese*, v. 67, n.2, pp. 249-58, 1986.
- FORSTER, Malcolm; SOBER, Elliott. “How to Tell When Simpler, More Unified, or Less Ad Hoc Theories Will Provide More Accurate Predictions.” *The British Journal for the Philosophy of Science*, v. 45, n. 1, pp. 1-35, 1994.

- GOODMAN, Nelson. *Fact, Fiction, and Forecast*. Cambridge, MA: Harvard University Press, 1955.
- GORDON, Emma C. “Understanding in Epistemology.” *The Internet Encyclopedia of Philosophy*, s/d.
- HASAN, Ali. “In Defense of Rationalism about Abductive Inference.” In: McCAIN, Kevin; POSTON, Ted. *Best Explanations: New Essays on the Inference to the Best Explanation*. Oxford, UK: Oxford University Press, 2017.
- HAWTHORNE, John. “Inductive Logic.” In: ZALTA, Edward N. (ed.). *The Stanford Encyclopedia of Philosophy*, 2018.
- HEMPEL, Carl G.; OPPENHEIM, Paul. “Studies in the Logic of Explanation.” *Philosophy of Science*, v. 15, pp. 135-75, 1948.
- HEMPEL, Carl G. *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*. New York, NY: The Free Press, 1965.
- HORWICH, Paul. *Probability and Evidence*. Cambridge, UK: Cambridge University Press, 1982.
- HOSSENFELDER, Sabine. *Lost in Math: How Beauty Leads Physics Astray*. New York, NY: Basic Books, 2018.
- HOWSON, Colin. “Exhuming the No Miracles Argument.” *Analysis*, v. 73, n. 2, pp. 205-11, 2013.
- JACKSON, Frank. “Grue.” *The Journal of Philosophy*, v. 75, n. 5, pp. 113-31, 1975.
- KUKLA, André. “Antirealist Explanations of the Success of Science.” In: *Philosophy of Science*, v. 63, n. 5, pp. S298-S305, 1996.
- LAKATOS, Imre. *Proofs and Refutations: The Logic of Mathematical Discovery*. Cambridge, UK: Cambridge University Press, 1976.
- LEWIS, David K. “New Work for a Theory of Universals.” *Australasian Journal of Philosophy*, v. 61, n. 4, pp. 343-77, 1983.
- LIPTON, Peter. *The Inference to the Best Explanation*. 2. ed. London, UK: Routledge, 2004 (1991).
- LIPTON, Peter. “Understanding Without Explanation.” In: DE REGT, Henk W. *et al* (eds.). *Scientific Understanding: Philosophical Perspectives*. Pittsburgh, PA: University of Pittsburgh Press, pp. 43-63, 2009.

- LOWE, E. Jonathan. “Ontological Dependence.” In: ZALTA, Edward N. (ed.). *The Stanford Encyclopedia of Philosophy*, 2015.
- LYCAN, William G. “Explanation and Epistemology.” In: MOSER, Paul K. (ed.). *The Oxford Handbook of Epistemology*. Oxford, UK: Oxford University Press, pp. 408-33, 2002.
- NOLAN, Daniel. “Hyperintensional metaphysics.” *Philosophical Studies*, v. 171, n. 1, pp. 149-60, 2013.
- PAUL, Laurie A.; HALL, Ned. *Causation: A User’s Guide*. Oxford, UK: Oxford University Press, 2013.
- PSILLOS, Stathis. “The Fine Structure of Inference to the Best Explanation.” *Philosophy and Phenomenological Research*, v. 74, n. 2, pp. 441-8, 2007.
- PSILLOS, Stathis. *Knowing the Structure of Nature: Essays on Realism and Explanation*. Hampshire, UK: Palgrave Macmillan, 2009.
- SALMON, Wesley C. *Four Decades of Scientific Explanation*. Pittsburgh, PA: University of Pittsburgh Press, 2006 (1989).
- SCHAFFER, Jonathan. “Grounding, Transitivity, and Contrastivity.” In: CORREIA, Fabrice; SCHNIEDER, Benjamin. *Metaphysical Grounding: Understanding the Structure of Reality*. Cambridge, UK: Cambridge University Press, pp. 122-38, 2012.
- SCHAFFER, Jonathan. “The Metaphysics of Causation.” In: ZALTA, Edward N. (ed.). *The Stanford Encyclopedia of Philosophy*, 2016.
- SCHINDLER, Samuel. *Theoretical Virtues in Science: Uncovering Reality through Theory*. Cambridge, UK: Cambridge University Press, 2018.
- SCHNIEDER, Benjamin. “A Logic for ‘Because’.” *The Review of Symbolic Logic*, v. 4, n. 3, pp. 445-65, 2011.
- SCHUPBACH, Jonah N. “Must the Scientific Realist be a Rationalist?” *Synthese*, v. 154, n. 2, pp. 329-34, 2007.
- SHER, Gila. “On the Explanatory Power of Truth in Logic.” *Philosophical Issues* (forthcoming).
- SKOW, Bradford. *Reasons Why*. Oxford, UK: Oxford University Press, 2016.
- SOBER, Elliott. *Simplicity*. Oxford, UK: Clarendon Press, 1975.
- SOBER, Elliott. “The Principle of Parsimony.” *British Journal for the Philosophy of Science*, v. 32, n. 2., pp. 145-156, 1981.

- SOBER, Elliott. “Explanation in Biology: Let’s Razor Ockham’s Razor.” *Royal Institute of Philosophy Supplement*, v. 23, pp. 73-93, 1990.
- SOBER, Elliott. “What is the Problem of Simplicity?” In: ZELLNER, A. et al. (eds.). *Simplicity, Inference and Modelling: Keeping It Sophisticatedly Simple*. Cambridge, UK: Cambridge University Press, pp. 13-31, 2001.
- SOBER, Elliott. *Ockham’s Razor: A User’s Manual*. Cambridge, UK: Cambridge University Press, 2015.
- STANFORD, Kyle. *Exceeding Our Grasp: Science, History, and the Problem of Unconceived Alternatives*. Oxford, UK: Oxford University Press, 2006.
- SWINBURNE, Richard. *Simplicity as Evidence of Truth*. Milwaukee, WI: Marquette University Press, 1997.
- THORDARSON, Sveinbjorn. *Simplicity as a Theoretical Virtue*. Unpublished essay. Available at: (https://sveinbjorn.org/simplicity_as_theoretical_virtue)
- VAN FRAASSEN, Bas. *The Scientific Image*. Oxford, UK: Clarendon Press, 1980.
- VAN FRAASSEN, Bas. *Laws and Symmetry*. Oxford, UK: Oxford University Press, 1989.
- WILLIAMSON, Timothy. “Abductive Philosophy.” *Philosophical Forum, Inc.*, v. 47, n. 3-4, pp. 263-80, 2016.
- WILSON, Jessica. “No Work for a Theory of Grounding.” *Inquiry: An Interdisciplinary Journal of Philosophy*, v. 57, n. 5-6, pp. 535–79, 2014.