Underdetermination: Can Inference to the Best Explanation provide a way out?

Bachelor's thesis for Philosophy

By Ernö Groeneweg, 4279662

Utrecht University, Humanities, Philosophy and Religious Studies

Theoretical philosophy

21-06-2018

Primary grader: Johannes Korbmacher

Secondary grader: Chris Meinz

Word count: 6540

CONTENTS

ABSTRACT

This thesis evaluates a specific argument within the debate between scientific realism and anti-realism. Specifically concerning a problem for scientific realism known as underdetermination, as put forward by Bas van Fraassen in *The Scientific Image:* the idea that empirical equivalence leads to inherent underdetermination of scientific theories, which therefore cannot be considered to be true. It is the aim of this thesis to defend the project of scientific realism against Van Fraassen's argument by taking a dive into various conceptions of the idea of Inference to the Best Explanation, which is a type of inference that claims that from the best explanation of a phenomenon the truth can be inferred. Three versions of IBE as put forward by Gilbert Harman, Richard Boyd and Peter Lipton are considered. Harman believes the best explanation can be found using Bayesian probability theory. Boyd argues for the validity of Inference to the Best Explanation on a much more empirical basis. Finally, Lipton details an incredibly concise account of Inference to the Best Explanation, defining it as Inference to the 'Loveliest' Explanation and detailing the various criteria that go with it. This final conception of Inference to the Best Explanation along with its criteria shows that the problem of Underdetermination need not be a problem at all, since empirical equivalency of contradicting theories does not necessarily entail underdetermination.

It seems that, as a society, we place a lot of weight on the achievements of the modern sciences. What science actually achieves regarding the external world, is the question that lies at the heart of the debate between scientific realism and anti-realism within the philosophy of science. A scientific realist regards science as giving us an approximately true story of the external world, whereas a scientific anti-realist would consider this to be impossible, or at the very least impractical. In this thesis I will outline a single but very important line of argumentation from the camp of anti-realism, specifically by Bas van Fraassen in his work *The Scientific Image* (Fraassen 1980), known as *the underdetermination problem*. This is roughly the idea that any scientific theory has an empirically equivalent theory, which is to say a theory which follows from the same body of empirical data but is inconsistent with the theory which science believes to be true. Therefore, every scientific theory is inherently underdetermined by the body of empirical evidence on which it rests. After defining underdetermination, I will try to defend the view of scientific realism against this argument using the writings of several scientific realists. I will focus on the idea of Inference to the Best Explanation. This is a form of inference which, instead of relying on more traditional logical forms of inference, allows one to infer from a set of premises to the best explanation for these premises. In this thesis, I aim to show that the argument concerning underdetermination per Van Fraassen is maybe not as serious a threat to the project of scientific realism as it might seem at first. First of all, it is important to understand the basic debate between scientific realism and anti-realism.

I. The Debate of Scientific (Anti-)realism

Scientific realists differ greatly in their approach, so it is hard to summarise what this view pertains in general terms. There does seem to be a central core of ideas. Typically, these ideas involve a positive epistemological attitude towards scientific investigation; science can bring us somehow closer to knowledge about external facts. In particular, we can attain this knowledge not only empirically, but also about things which are not directly observable. It is important to understand the three general theses which in conjunction form the general idea of scientific realism, at least in the sense that I will understand it for the sake of the current argument. These three theses are ontic, semantic and epistemic realism (Psillos 1999, xix). Ontic realism is roughly the idea that the objects that scientific theories seem to posit actually exist, semantic realism is roughly the idea that our most successful scientific theories at this point are true or at least approximately true, and epistemic realism is characterised by any positive attitude towards the possibility of attaining knowledge about reality. To illustrate this point, consider electrons. Electrons cannot be directly observed. However, scientific realism would hold that electrons actually do exist, or at least that our idea of what electrons are, is approximately equivalent to the way they actually are. Our idea of electrons is not some abstraction

or useful fiction, nor can we simply consider it to be an instrument to explain our interpretation or phenomenology of the world. The scientific theories which posit the existence of electrons, are also considered to be (approximately) true; when I describe the current through a wire as electrons traveling from one end to another, I am giving a *literally true* account of what actually happens. This is in line with Van Fraassen's interpretation of scientific realism, which is what he directly argues against. According to Van Fraassen, the scientific realist would argue that science aims to understand the literal truth, which is to say a literal description of our external world. In this view, all that matters for accepting a scientific theory is the belief that it is literally true, which is to say that it gives a literal description of something within that external world. (Fraassen 1980, 8). Again, if we return to our idea of electrons: if we accept electrons to be the way that theoretical physics tells us they are, then we believe that that way is literally the way they are. The opponents of scientific realists are naturally called scientific anti-realists. Since scientific anti-realism is the negation of scientific realism, it can be considered the disjunction of the opposite positions that make up scientific realism: ontic, semantic and epistemic anti-realism. Ontic anti-realism within this debate is the view that the objects that are postulated by scientific theories cannot be considered to actually exist, semantic anti-realism is roughly a position wherein one believes our best scientific theories cannot be considered to detail some true story of the external world, and epistemic anti-realism is the view that even if ontic and semantic realism would hold, we still would have no way of knowing anything for sure. A very influential variant of scientific anti-realism is put forth by Bas van Fraassen in *The Scientific Image* (Fraassen 1980). His variant of scientific anti-realism is a direct negation of his own slightly different view of scientific realism. According to him, "anti-realism is a position according to which the aim of science can well be served without giving (…) a literally true story" (Fraassen 1980, 9). It seems to be an attempt to grant science as many virtues as he can, without having to account for the semantic and ontic commitments that scientific realists embrace. Generally, when a scientist proposes a theory, we tend to think that the scientist is asserting it to be true. A scientific anti-realist would instead state that he simply displays it and proclaims certain virtues of it… but no notion of truth. "These virtues may fall short of truth: empirical adequacy, perhaps; comprehensiveness, acceptability for various purposes" (Fraassen 1980, 10). Van Fraassen calls his particular brand of scientific anti-realism *constructive empiricism.* It's the idea that scientific theories are empirically adequate, and that the belief of a theory involves *only* the belief that the theory is empirically adequate (Fraassen 1980, 12). Per Van Fraassen, a theory is *empirically adequate* exactly if what it says about *observable* things in the world is true. To return to our example of the electron once more: what we can directly observe of electrons is that if we connect wires between a power source and a lightbulb, the lightbulb turns

on. If this is all our scientific theory states to be true, then it is empirically adequate. Then, if we believe the theory, we believe that it is empirically adequate and nothing more.

II. Empirical Equivalency and Underdetermination

In his project to reinvigorate scientific anti-realism, Van Fraassen asks: what is the empirical content of a scientific theory? (Fraassen 1980, 41) He investigates the connection between scientific theorisation and the empirical data on which it rests. Van Fraassen analyses a specific example: Newton's theory of planetary motions. Newton was the first to postulate that there was such a thing as 'absolute space' and that the motion of all bodies is relative to it as well as relative to one another. The absolute motions for Newton were part of the axioms, which could be calculated from empirical data of the movement of bodies relevant to earth. In his model, the centre of gravity of our solar system is at rest with absolute space, but because of the nature of relative movement, the centre of gravity might as well have been at any other magnitude of absolute motion. This means that for every hypothesis which differs from Newton's on an axiomatic level—that is, with another absolute motion for the centre of gravity of the solar system—is still *empirically equivalent*—that is, they do not differ in their empirical aspects. Van Fraassen names Leibniz' cosmology, where space has a certain constant velocity in a certain direction. Leibniz' hypothesis would be empirically equivalent to Newton's, since they follow from the same empirical data, yet they both consider something to be true about the universe which is in contradiction one another's hypothesis. (Fraassen 1980, 46-47). This idea of empirical equivalency leads to the problem of underdetermination that scientific realism has to face.

Generally, the problem of underdetermination is formulated as follows: consider a language to be the union between a set of observational sentences O and a set of hypothetical sentences H: $L = O \cup H$. Under plausible assumptions, there is for every scientific theory T a theory T', for which two things hold:

1) T and T' are *empirically equivalent*, which means that their axioms are inferred to from the same empirical data and thus that they say are equivalent within O, which also means that they are both empirically adequate.
2) T and T' are in contradiction somewhere in H.

Since the axioms which were inferred to via empirical means are all that scientific theories are based on, we can never know whether our best scientific theories are approximately true, nor whether all the entities that these theories postulate actually exist. In this way, the idea of empirical equivalency leads to the assumption that scientific theories will always be *underdetermined* by the empirical data from which they have been inferred. Whatever the two theories might agree on in terms of the

observational language, on any other front we can never be certain they describe the same thing; we can never be certain one is closer to the truth than the other (Fraassen 1980, 47).

III. Inference to the Best Explanation

An attempt to rescue the project of scientific realism from the problem of underdetermination can hopefully be found in the idea of *Inference to the Best Explanation* (IBE). The general idea here is that within science, there is a hypothesis which is the 'best explanation' of the data it is based on. IBE allows us to infer that the best explanation for our observations is (approximately) true. More formally: consider two hypotheses H and H' that both explain some set of observations O. We can infer that H is true rather than H' precisely if H is a *better explanation* for O than H' based on certain criteria (Fraassen 1980, 19). In order to conclude whether a version of IBE might find a way out of Van Fraassen's problem of underdetermination, we need to examine four authors who have put slightly different version of this idea forward. The first version of this form of inference to examine is the one described by Gilbert Harman.

1. Harman

According to Harman in *The Inference to the Best Explanation* (G. H. Harman 1965), IBE corresponds approximately to what "others have called 'abduction,' 'the method of hypothesis,' 'hypothetic inference,' 'the method of elimination,' 'eliminative induction,' and 'theoretical inference.'" (G. H. Harman 1965, 88) This should already give some insight into what is meant by the term: it obviously has something to do with hypotheses. By making an inference to the best explanation, one is inferring from the fact that a certain hypothesis would explain the evidence to the truth of that hypothesis. This is a fairly standard reconstruction. For Harman however, all inductive inference takes the form of IBE: when we use induction, we generally make some form of generalisation based on a limited amount of empirical data. This generalisation is considered the 'best explanation' for the data in consideration (G. H. Harman 1965, 88). One argument in favour of using IBE for Harman is just how widespread the usage of IBE actually is. Not just in science, but in detective work, witness interrogation, a court of law and many other everyday scenarios IBE is used all the time. It is a pattern of inference that is sometimes considered the most common of all types of inference. We seem to follow the rule of IBE constantly (G. H. Harman 1965, 89). A question that Harman needs to answer, of course, is by what criteria one might determine which possible hypothesis gives the best explanation. He mostly delves into the idea of evaluating an explanation based on Bayesian *unconditional probabilities*, which is the idea that the probability of something being true can be found *a priori,* based on various factors instead of the idea that a probability is always based on empirical

data, frequencies or something in that regard (Joyce 2016). H is a better explanation than H' of E, provided that:

(a) P(H) > P(H')—the probability of H being larger than the probability of H'—and

(b) P(E|H) > P(E|H')—the probability of E given H is larger than the probability of E given H'.

This is also popularly called IPE, or *Inference to the most Probable Explanation.* According to Harman, this is a consideration people make constantly. If you are flipping a coin, for instance, of which you assume it is a normal balanced and 'fair' coin. Consider this hypothesis H. The probability of this is of course far larger than the probability of the coin being weighted, which we could call hypothesis H'. You flip it 20 times, 15 times of which it lands on tails and only 5 times it lands on head. Consider these observations our 'E'. Suddenly, the hypothesis that the coin is weighted seems more probable, given the fact that the probability of the coin landing 3 times as often on tails instead of heads is larger if the coin is weighted than if it is not. Now, it is still a lot more probable that the coin is still balanced and that itis just a coincidence that it landed on tails more often. In practice, one would test this out by flipping the coin many times more. This of course changes the evidence on which we base our probabilities in step (b). The bigger the difference between P(E|H) and P(E|H') becomes, the more likely it becomes that H' is in fact a better explanation of E than H. (G. Harman 1968, 169-170)

As a solution of the problem of underdetermination, this version of IBE does not hold. This is simply the case because Van Fraassen has written a convincing rebuttal to this version of IBE himself (Fraassen 1980, 19-23). One problem is the use of these Bayesian probabilities; step (a) in the formal description of the IBE criterium supposes some *a priori* inherent probability of H and H', whereas traditional statistical practice only deals with probabilities of something given some other thing. A traditional Bayesian might try to solve this problem by stating that we all ascribe some subjective probability to everything, so the necessary probabilities are theoretically all available. By making these probabilities subjective, however, much of the strength of IBE is lost. We need hard criteria to determine which explanation would be the best, which we cannot do if those criteria rely on the subjectivity of things like Bayesian probabilities (Fraassen 1980, 22). Another criticism against IPE is the fact that it does not actually tell us anything about the truth of the explanation in question. There is no clear connection between the probability of an explanation and whether it is actually true or not. A not very *probable* explanation is still very much *possible*. Why should we believe that an explanation is true simply because it is the most probable one? Since there is still the very real possibility of another explanation, the explanation is still underdetermined. It seems clear that inference to the most probable explanation does not really seem to find a way out of the problem of underdetermination. Let's see if another account can find us a way out.

2. Boyd

Richard Boyd has developed a well-known empirical argument in favour of IBE, mostly in *Scientific Realism and Naturalistic Epistemology* (Boyd 1981). Boyd does not define IBE much differently than the standard way, but he does offer an interesting way out of the idea of empirical equivalency. Boyd argues that scientific methodology is heavily theory-dependent; when scientists develop an experimental setup, or hypothesise about an explanation for certain phenomena, they draw heavily on other established scientific theories which have by and large been inferred to using some form of IBE. Since the sciences seem so heavily theory-dependent and yet have often yielded and continue to yield incredibly accurate results, Boyd concludes that we can do little else than believe that IBE does in fact bring us close to at least approximate truth (Boyd 1981, 615-616). A popular criticism against this line of argumentation is that there seems to be some circularity to it. Clearly, one of the premises of the argument is that scientific methodology is informed by a body established scientific theories, which are considered approximately true. This is a premise that rests on IBE itself to be considered plausible; it is the best explanation for the apparent success of science that established scientific theories are approximately true. Yet it is the validity of IBE itself that Boyd is trying to infer to (Laudan 1981, 20-22). Before we can even consider this approach as a way out of the problem of underdetermination, we must first see if this circularity is actually present.

For a solution, we need look no further than Stathis Psillos' work *Scientific Realism: How Science Tracks Truth* (Psillos 1999). Psillos asks whether Boyd's argument can actually be considered so 'viciously circular'. To investigate this, he differentiates between two types of circularity: premise-circularity and rule-circularity; an argument is premise-circular precisely if the conclusion is also one of the premises, whereas within a rule-circular argument one of the premises is an assertion about an inferential rule that is itself used to infer the conclusion of said argument (Psillos 1999, 82). According to Psillos, Boyd's argument is *not* premise-circular, which is always *viciously* circular and therefore considered a fallacy, but it *is* rule-circular, which does not need to be viciously circular. In Psillos' view, an argument that argues in favour of the reliability of a certain rule but relies itself on that rule, is not fallacious as long as the use of that rule does not guarantee a positive conclusion about the reliability of the rule. Psillos argues that although Boyd's argument for the reliability of IBE does use IBE to argue that scientific theories are probably approximately true because they are so successful, this use does not guarantee the *truth* of that conclusion. In fact, all that Boyd says is that the idea that established scientific theories are approximately true is the best explanation for how successful they are, and that is all. Because of this, Psillos believes Boyd's argument to be sound (Psillos 1999, 85-86).

We do however need to make our own assessment of Boyd's argument. We may accept it to be sound, but does it offer any way out of the problem of underdetermination as resultant from

empirical equivalency? It does seem to hint at the possibility, by stating that in forming a hypothesis, scientists do not solely rely on empirical data. Not only do they infer to a hypothesis as the best explanation for the given data *as well as* established scientific knowledge, they also take this established knowledge into consideration when designing the experiments they use to attain the empirical data in the first place. This way, empirically equivalent hypotheses do no longer necessarily show underdetermination, since they also both need to agree with the existing body of established scientific knowledge. This would at the very least make it a lot more challenging to define an empirically equivalent hypothesis that also seems to suggest some form of underdetermination. We may be going in the right direction here, but there is still a lot left unsaid about the actual methodology of IBE. It might *in principle* offer a way out, but there is always the question of: how does it actually work? Is it even usable? If we cannot answer these questions, then we are still not warranted to believe that our best scientific theories are not underdetermined by their empirical data, much less that they are approximately true. To figure this out, we look to Peter Lipton.

3. Lipton

Peter Lipton has provided us with a very comprehensive and influential work on IBE, aptly named *Inference to the best Explanation* (Lipton 1993). There are two important distinctions to be made when examining Lipton's version of IBE: the distinction between the actual and all potential explanations, and the distinction between the explanation best supported by the evidence and the explanation which would provide the best understanding: the distinction between the *likeliest* and the *loveliest* explanation (Lipton 1993, 59).

Inference to the Loveliest Explanation

The distinction between the actual explanations and potential explanations is necessary for Lipton because of three reasons (Lipton 1993, 59). Firstly, IBE cannot be understood as inference to the best *actual* explanation, since every 'actual explanation' can be assumed to be approximately true. Of course, we must allow reasonably that one can infer to a falsehood. Secondly, we can hardly suppose competing explanations—that are often incompatible! —to all be *actual*. In the context of the problem of underdetermination, this would commit us to saying that both empirically equivalent hypotheses are approximately true, which they obviously are not since they are by definition of the problem incompatible with one another. Thirdly and most importantly, we cannot infer to the best *actual* explanation simply because we have no way of knowing which explanations are 'actual' until after we have made the inference. We want to explain the connection between empirical evidence and inference, but by wanting to infer to the best actual explanation we presuppose that very connection. So, we must allow *potential* explanations to play a role in our inferential process, and we must distinguish between potential and actual explanations. Then, we can define IBE broadly as Inference

to the Best *Possible* Explanation: we have a pool of possible explanations and we infer to the best one in there, which is an *actual* explanation. Now we can allow unsuccessful inferences, different explanations can be incongruous because they are only *potential* and we do not need to presuppose a notion of truth before we even start inferring (Lipton 1993, 59-60). A pool of all potential explanations can become very large, so Lipton wants to narrow this selection down to only those explanations that are *plausible*, that realistically might become actual explanations. This selection process is part of the practice of IBE according to Lipton. When we infer to the best explanation, we first filter out all explanations that are implausible, and then we select the best explanation from the set of explanations that is left (Lipton 1993, 61).

The second distinction that Lipton makes is between what he calls the 'likeliest' and the 'loveliest' explanation. The 'likeliest' explanation is the explanation that is the most warranted given the evidence that we have. The 'loveliest' explanation is the explanation that is the most 'explanatory'; it provides the most understanding. The question then becomes which of these characterisations would fit the label of 'best' explanation, or in other words: which criterion will we consider the main criterion for selecting the best explanation in our inference? Important for the distinction between the 'likeliest' and the 'loveliest'-criteria is that they often result in very different explanations, depending on which criterion you choose to follow (Lipton 1993, 61-62). The likeliest explanation is not always very lovely and vice versa. The main point of divergence between the two, according to Lipton, is that likeliness is relative to the *total* balance of evidence, whereas loveliness does not necessarily have to be. Lipton draws on Newtonian physics to illustrate his point (Lipton 1993, 62). It is probably still one of the loveliest explanations for the way bodies interact in physics, but because of the advent of special relativity and quantum mechanics it is becoming less and less *likely*. It is far easier to understand how bodies interact with one another when viewing it through the lens of Newtonian mechanics, but there is some evidence that is ignored or at least not given a lot of importance in Newtonian physics. On the other hand, quantum mechanics and special relativity are very difficult to understand, but at least these take all available evidence into consideration in equal measure. Generally, the advent of new potential explanations might decrease the *likeliness* of an older explanation, but it need not decrease its *loveliness* (Lipton 1993, 62). When choosing one over the other, the likeliness criterion begins to fall short. The project of IBE is to give a model of inductive inference that describes by what principles one explanation is *more likely* than another. By saying that the most likely explanation is the one that is more likely than another, we are of course begging the question. This is the reason that Lipton opts in favour of IBE defined as *Inference to the Loveliest Potential Explanation* (ILPE). By this definition, the explanation which provides the most understanding is most likely to be true. (Lipton 1993, 63).

The Criteria for Loveliness

The obvious next question that arises is: how do we determine which explanation is the loveliest? For this, Lipton has several criteria. We will consider them one by one, thus trying to show how ILPE can in fact be a good way to avoid the problem of underdetermination. There is the *Difference Condition,* which looks into differences in the causal history of phenomena. The *Mechanical Condition* shows that the causal mechanism at play between phenomena and their causal histories play a role in our understanding. The *Precision Condition* tells us that the more precise explanation is lovelier, and the *Unification Condition* requires that a lovelier explanation is more easily unified with existing data.

Let's begin by looking at Lipton's causal model of explanation, which is the foundation of the Difference Condition. In the causal model of explanation, to explain a phenomenon one need only give information about its causal history, or to explain it is to "give information about the mechanisms linking cause and effect", in the case of a causal regularity (Lipton 1993, 32). How do we then determine which pieces of the causal history of a phenomenon serve as explanatory and which do not? According to Lipton, this can be determined by contrasting. The example he gives about his three year old son is apt: when the child throws his food on the floor, you ask him why and he replies that he did it because he was full, then that might explain why he threw it on the floor rather than eating it, but it is not an explanation as to why he threw it on the floor instead of leaving it on his plate (Lipton 1993, 35-36). When asking for a good explanation, one should ask "why P rather than Q?" or, as Lipton calls it: "why the fact rather than the foil?" (Lipton 1993, 36). We presuppose the fact occurred and the foil did not, but the *reason why* the fact occurred and not the foil is in fact the explanation for the phenomenon. It's the contrast between the fact and the foil where the explanation can be found. This is, of course, perfectly in line with the traditional conception of IBE; the explanation closes the gap and explains why we should choose for H rather than H'. Lipton calls this contrastive version of explanation in terms of the causal model the *Difference Condition:* "To explain why P rather than Q, we must cite a causal difference between P and not-Q, consisting of a cause of P and the absence of a corresponding event in the case of not-Q" (Lipton 1993, 43). Say you have two identical matches. The only thing that's different between the two is that match M is dry and match M' is not. If M lights and M' does not, and the only difference between the two is their dryness, then the fact that M' is not-dry is the explanation why it did not light (fact) rather than that it did (foil) (Barnes 1995, 253-254).

For Lipton, the *Difference Condition* is one important criterion for determining the degree of loveliness of a potential explanation (Lipton 1993, 75-76). Consider two empirically equivalent and contradicting theories T and T'. Theory T is an explanation which allows for a greater understanding as to why some phenomenon P and not some other phenomenon Q is the case, with regards to the

causal history of P and Q. It is a lovely explanation if it shows the difference in the causal histories of P and not-Q which has resulted in the fact that P and not-Q came to pass. On the other hand, T' is an explanation which makes this fact-foil difference less clear. A problem arises however when there are multiple differences between the fact and the foil. Then, the difference condition does not grant an explanation with a good level of understanding, which of course makes it a lot less lovely. Does this offer a way out of the problem of underdetermination? Maybe not on its own, but it does mean that in order to formulate an explanation, one needs to look beyond just the direct empirical evidence that needs to be explained: the causal history of all the evidence is also taken into consideration, which greatly restricts the amount of lovely empirically equivalent explanations and gives us some tools to measure the loveliness of each of those explanations. Beyond this, we have of course many more criteria to examine.

Barnes in *Inference to the Loveliest Explanation* (Barnes 1995) believes that Lipton endorses something called the *Mechanism Criterion*, which he summarises in this way: "an explanation is more lovely the more fully it describes the mechanism by which a putative cause brings about its effect" (Barnes 1995, 257). The example given here is apt. The empirical data we have is that Harry hit the dog Fido, rather than another dog named Bozo. In order to find the best explanation for this event, we look at the causal history of both P (Harry hitting Fido) and not-Q (Harry not hitting Bozo). We consider two differences between in these histories: Fido bit Harry but Bozo did not. Let's call this the Bite Hypothesis. Another difference is that Bozo reminded Harry of a dog he had as a child while Fido did not. Let's call this the Childhood Hypothesis. Most people would agree that the fact that Fido bit Harry is a better explanation for Harry hitting Fido than the fact that Fido did not remind Harry of his childhood dog but Bozo did. This is because there is a familiar mechanism at play: when you are injured, you are more likely to become angry or fearful and strike back. Being reminded of one's childhood dog and therefore hitting another dog is not a very good explanation. That would be an explanation that would offer less of an understanding as to why Harry hit Fido (Barnes 1995, 258). The Mechanism Criterion can be read as an extension of the Difference Criterion, allowing us to measure the loveliness of empirically equivalent theories even more strictly. The possibility of creating empirically equivalent theories will shrink more and more, the more precisely we can differentiate between theories and start to find differences between them that could very well help us discover which explanation is the loveliest.

Another criterion that Lipton proposes for explanatory loveliness is the *Precision Criterion*. An explanation is more lovely the more precisely the phenomenon is explained by the explanation (Barnes 1995, 260) (Lipton 1993, 118). Consider the aforementioned Bite Hypothesis and the Childhood Hypothesis once more. The fact that Fido bit Harry is a more precise explanation for the

fact that Harry hit Fido and not Bozo, simply because thanks to the mechanism described above, being bitten by a dog entails a violent reaction, more so than the fact that the dog did not remind you of your childhood dog. The explanation that Fido did not remind Harry of his childhood dog, could point to many other phenomena happening rather than Harry hitting Fido, whereas Fido biting Harry more precisely explains the phenomenon of Harry hitting Fido. It may be said that it is still possible to create two empirically equivalent theories that also explain the set of empirical data equally precisely, but it is yet another criterion that makes such a formulation increasingly difficult.

Another criterion that Lipton considers is the *Unification Criterion*. At its core, this criterion states that a lovely explanation does not posit any other phenomenon that cannot be explained within our conception of the world. The example given is the 'sympathetic powder' from Sir Kenelm Digby, in the 17th century. Sir Digby hypothesised that a sword cut could be cured by rubbing a 'sympathetic powder' on the offending sword, rather than trying to treat the wound. The evidence for this? Victims of sword cuts treated in ways that were regular for the time, recovered significantly slower than victims that were not treated directly. Of course, the currently accepted explanation is simply that established treatments at the time did more harm than good, and simply leaving the wound alone was more beneficial to the healing process. Why is this a more lovely explanation for the phenomenon than the fact that the 'sympathetic powder' actually worked? According to Lipton, this is simply because the sympathetic powder-explanation would require accepting some underlying phenomenon that shows it's possible to affect something over a distance when rubbing powder on a sword, whereas the accepted explanation can be more easily *unified* with our existing knowledge and conception of the world (Barnes 1995, 262) (Lipton 1993, 118). In this way, ILPE relies not just on the body of evidence it is directly inferred from, but on all accepted knowledge and understanding about our world. It thus becomes a lot more difficult to create a theory T' which is empirically equivalent yet contradictory for a theory T which follows the unification criterion. Not only do the two theories agree observationally, but they also need to agree with other theoretical knowledge. This means that even empirically equivalent theories that could contradict in any other subset of language must be measured against matters of fact within other subsets of language themselves[1]. Therefore, empirical equivalence alone does not mean that we have no way of knowing which theory is closer to the truth.

The final criterion of unification brings the validation of explanations outside of its confines within the realm of observational language, and thereby seems to also turn it away from empirical equivalency being a serious problem. *Inference to the Loveliest Explanation* tells us that the best explanation is the explanation which provides us the greatest understanding of the phenomena it explains. We can determine which explanation this is by following various criteria, such as asking

[1] Recall how I defined 'empirical equivalence' in part as "agreeing in the observational subset of language".

contrastive questions, looking at the underlying mechanisms of these contrasts, inquiring after the precision of an explanation and, arguably most importantly, always taking other knowledge (theoretical or otherwise) into account and attempting unification with that knowledge. This of course sounds similar to what Boyd attempted to do in his defence of IBE as a valid and approximate-truth-tracking part of scientific methodology. However, rather than relying on the empirical claim that IBE can attain approximate truth because scientific methodology is so accurate, Lipton decided instead to develop a very precise and thorough account of IBE that, instead of relying on empirical fact, rather shows a system of logical inference that is not only likely widely used already but also provides some truth-tracking merit to scientific methodology. By precisely showing us the methodology behind ILPE and showing how hypotheses do not rely solely on the empirical data which they explain, Lipton has shown a very promising way out of the problem of underdetermination.

The solution of IBE

I have detailed three distinct accounts from three different philosophers concerning the methodology of Inference to the Best Explanation. We saw that Harman's project of using Bayesian probability theory to infer to the best explanation did not yield favourable results, as Van Fraassen himself pointed out. Next, we looked at Boyd's definition, which sought to argue in favour of IBE as a valid and sound method of inference within science from an empirical standpoint. Although his argument seemed circular, Psillos showed us that this circularity does not need to be a problem. Although Boyd claimed that an inference to the best explanation takes more into account than only the empirical data of the problem at hand, he failed to show us how it would work exactly. This is where Lipton came to the rescue. By not only defining IBE as 'inference to the loveliest explanation'—where the loveliest explanation is the explanation which grants the most understanding—but also showing how the loveliness of an explanation can be judged using a variety of methods, a possible concrete methodology for IBE became apparent. Boyd has shown that when inferring to a hypothesis from empirical data, more things beyond the strictly empirical have been considered. If we combine this idea with the clear and concise methodology that Lipton has provided, we can show that empirically equivalent theories do not necessarily follow equally well from the available empirical data. Because this is no longer necessarily true, we can confidently state that underdetermination which follows from empirical equivalency is not a problem anymore. Combining this with Peter Lipton's detailed methodology of the IBE, we have shown that scientific theories can in fact be considered to be (approximately) true, even when empirically equivalent theories exist.

Conclusion

I have shown that the principle of *Inference to the Best Explanation* has some merit when it comes to finding a way out of the problem of underdetermination as posited by Bas van Fraassen. Gilbert

Harman's original account of the method was lacking in various areas and his dependency on Bayesian probability theory left him wide open to counterarguments from Van Fraassen himself. Richard Boyd showed a very simple yet powerful argument in favour of IBE as an integral and properly truth-tracking component of scientific methodology. However, his argumentation was shaky at best, despite Psillos' best efforts to defend its apparent circularity. On top of that, it did not tell us much about how IBE would actually work in practice, leaving us open to the criticism of defending an ultimately unusable method of inference. That is why we turned to Peter Lipton's influential work *Inference to the Best Explanation,* which showed an extremely thorough and carefully constructed version of IBE as *Inference to the Loveliest Explanation*. By carefully demarcating what an explanation actually is, how we can arrive to an explanation and listing clear criteria as to how an explanation can be evaluated to be the 'best', Lipton has made it considerably more difficult for the scientific anti-realist to keep insisting on underdetermination as empirical equivalency as a serious problem for scientific realism. There may still be angles for the scientific anti-realist to take and many other conceptions of underdetermination to consider, but the version that relies on empirical equivalency simply does not seem to hold up against IBE as a careful, truth-tracking and not solely empirically-dependent part of scientific methodology.

WORKS CITED

Barnes, Eric. 1995. “Inference to the Loveliest Explanation.” *Synthese* (Springer) 103 (2): 251-277.

Boyd, Richard. 1981. “Scientific Realism and Naturalistic Epistemology.” Edited by P. Asquith and R. Giere. *PSA 1980* (Philosophy of Science Association) II: 613-662.

Fraassen, Bas Van. 1980. *The Scientific Image.* Oxford: Oxford University Press.

Harman, Gilbert H. 1965. “The Inference to the Best Explanation.” *The Philosophical Review* 74 (1): 88-95.

Harman, Gilbert. 1968. “Knowledge, Inference, and Explanation.” *Amarican Philosophical Quarterly* 5 (3): 164-173.

Joyce, James. 2016. *Bayes' Theorem.* Edited by Edward N. Zalta. Accessed 06 14, 2018. http://plato.stanford.edu/archives/win2016/entries/bayes-theorem/.

Laudan, Larry. 1981. “A Confutation of Convergent Realism.” *Philosophy of Science* (The University of Chicago Press) 48 (1): 19-49.

Lipton, Peter. 1993. *Inference to the best Explanation.* New York: Routledge.

Psillos, Stathis. 1999. *Scientific Realism: How Science Tracks Truth.* New York: Routledge.