Scientific Revolutions and the Explosion of Scientific Evidence[1,2]

Ludwig Fahrbach
to appear in *SYNTHESE*

ABSTRACT Scientific realism, the position that successful theories are likely to be approximately true, is threatened by the pessimistic induction according to which the history of science is full of successful, but false theories. I aim to defend scientific realism against the pessimistic induction. My main thesis is that our current best theories each enjoy a very high degree of predictive success, far higher than was enjoyed by any of the refuted theories. I support this thesis by showing that the amount and quality of scientific evidence has increased enormously in the recent past, resulting in a big boost of success for the best theories.

1 Introduction

Scientific realism, the position that successful scientific theories are likely approximately true, is the dominant view among scientists, philosophers, and the educated public. However it faces formidable challenges, the most serious of which is often considered to be the pessimistic induction. The pessimistic induction is based on the observation that the history of science is full of theory changes or "scientific revolutions", in which once successful and accepted theories were refuted and abandoned. These refutations represent counterexamples to scientific realism and threaten to undermine it.

Realists have typically reacted to the pessimistic induction by substantially weakening their position: Successful theories, including our current best theories, can no longer be accepted as approximately true, but only as partially true. Different authors cash out the partial truth of a theory in different ways. Some take this to be the referential success of its theoretical terms, others the truth of its structural content, or the parts that "contribute to success", or still other parts[3]. Much effort has then been invested into showing that in most instances of past theory change the parts of the refuted theories constituting their partial truth survived the theory changes.

In this paper I show that realism can be defended against the pessimistic induction without having to regard our current best theories as only partially true. At the core of my defence are the following claims. Over the last few decades there has been a huge increase in the quantity, quality and diversity of scientific evidence. This increase has contributed strongly to the predictive success of many recent scientific theories. Practically none of these theories have been refuted. Therefore, and this is the main thesis of the paper, our current best theories enjoy a degree of predictive success far higher than those enjoyed by any refuted theories, practically all of which enjoyed moderate degrees of success at best, before they were refuted. Hence, if we restrict realism to theories with levels of success typical of our current best theories, it is not undermined by the pessimistic induction.[4]

[1] I would like to thank Michael Anacker, Luc Bovens, Hasok Chang, Michael Devitt, Brigitte Falkenburg, Bernward Gesang, Stephan Hartmann, Paul Hoyningen-Huene, Felicitas Krämer, James Nguyen, Helmut Pulte, Eric Schliesser, Gerhard Schurz, Mark Siebel, several anonymous referees, and audiences in Konstanz, Manchester, Duesseldorf, Dublin, Oldenburg, Bochum, Dortmund, Toronto, and Bern. For detailed comments on this or earlier drafts I would like to thank Claus Beisbart, Sungbae Park, Sam Ruhmkorff, Ioannis Votsis, and Paul Thorn. It is highly probable that I forgot some people, for which I apologize.

[2] The final publication is available at Springer via http://dx.doi.org/10.1007/s11229-016-1193-y.

[3] See Richard Boyd (1990), Ian Hacking (1983), John Worrall (1989), Philip Kitcher (1993), Martin Carrier (1993), Jarrett Leplin (1997), Stathis Psillos (1999), Steven French and James Ladyman (2003)

[4] Similar claims and arguments are presented by Gerald Doppelt (2007), Sungbae Park (2011), and Ludwig Fahrbach (2009, 2011a, 2011b). My paper builds on their work. It aims to fill several gaps in the argumentation of those approaches and pro-

I will proceed as follows. In section 2.1, I define scientific realism and introduce the no-miracles argument. In section 2.2 I introduce the pessimistic induction. The focus in these two sub-sections is on the class of *all* successful theories, past, present, and future. Then, in response to the pessimistic induction, I restrict realism to theories whose empirical success matches (or betters) the empirical success enjoyed by our current best theories. This move faces two objections, which I call the objection from the past and the objection from ad-hocness (section 2.3). The objection from the past states that people in the past could have restricted realism in an analogous way, namely to those theories whose degree of empirical success is the same (or higher) than that enjoyed by the best theories *of their time*, but they went on to be refuted nonetheless. The objection from ad-hocness states that restricting realism to current levels of success is ad hoc. I rebut both objections with the help of the main thesis (section 5.2 and 5.3). The argument for the main thesis is developed in sections 3, 4, and 5.1. Section 6 provides a conclusion.

2 Scientific realism and the pessimistic induction

2.1 Scientific realism and the NMA

I start with a common definition of scientific realism, which I modify later on. According to this definition, realism states that a successful theory is likely to be approximately true. We can assume that this definition is equivalent to the claim that the inductive inference from the success of a theory to its approximate truth is cogent.[5] I identify success with predictive success: A theory counts as successful just in case it has made sufficiently many true predictions of sufficiently high quality, where the predictions of a theory are its known observable consequences. If a theory makes false predictions, it does not count as successful. I use the terms “success”, “predictive success”, and “empirical success” interchangeably.

When realists talk about successful theories, the ones they have in mind are our current best theories, for example, the atomic theory of matter, the theory of evolution, and the germ theory of disease. These theories are undeniably highly successful, therefore realism implies that they are likely to be approximately true. The degree of truth approximation may depend on the respective scientific theory or scientific field, but I assume it to be generally fairly high.[6] In particular, a theory that is only partially true, e.g., only true about entities, structure or something else, does not count as approximately true. For the sake of brevity, I will sometimes omit the terms “likely” and “approximately”. Then realism simply states that successful theories are true.

Realists support their position with the no-miracles argument (NMA). In its simplest form the NMA states that if a successful theory were false, then the success of the theory would be a miracle (Putnam 1975). For example, if infectious diseases behaved on numerous occasions as if they were caused by microbes, viruses and parasites, but they are actually caused by something else entirely, then the success of the germ theory of disease would be a miracle. This version of the NMA appeals directly to our confirmational intuitions. Other more elaborate versions, which I will not use in this paper, rely on novel predictions or are connected with inference to the best explanation or similar kinds of reasoning (Boyd 1983, Alan Musgrave 1988, Psillos 1999). Anti-realists reject the NMA, of course. However, defending it is not an aim of this paper, rather I will simply assume that it has probative force, – although this does not preclude the possibility that it is trumped or refuted by stronger arguments, as we will see in a moment.

The realism debate has grown to epic proportions. There are many versions of realism, anti-realism, the NMA and the pessimistic induction. When attempting to construct a path through this thicket of extant positions and arguments one faces a combinatorial explosion of possibilities. Difficult choices about which views and considerations to discuss and which to ignore are unavoidable. My overall aim in this paper is limited, namely to formulate *one* viable path through the thicket. For example, I will not engage with Psillos’ distinction between first order and second order evidence

vide a more solid foundation for the whole dialectic.

[5] The term “inductive inference” is meant to cover all valid non-deductive inferences.

[6] On truth approximation, see, e.g., Ilkka Niiniluoto (1999) and Theo Kuipers (2000).

(2009). Furthermore, I won't deal with constructive empiricism, but will assume that the NMA has probative force. As a consequence, my defence of realism against the pessimistic induction can only claim to be an internal defence of realism. However, I will make it clear which parts of the argumentation rely on the NMA, and which don't. Moreover, I only use the simplest form of the NMA. I believe that using more elaborate versions would not change the dialectic significantly, but cannot show that here. Further choices and assumptions will be noted later on. Finally, the main claims of the paper are very general statements about the past development of science. Hence, parts of the argument have to remain sketchy and programmatic. A lot of work has to be left for other occasions, including the discussion of many objections.

2.2 The pessimistic induction

So, let's start with the dialectic. Realism is threatened by the pessimistic induction (PI) according to which the history of science is full of theories that were once successful and accepted by scientists as true, but later refuted and abandoned. A long list of such theories was famously presented by Larry Laudan (1981). It includes, for example, the phlogiston theory, the caloric theory, and the ether theory.[7] Thus, realists face a situation in which there are two arguments pulling in opposing directions: The NMA supports realism, and the PI undermines realism. The two have to be balanced against each other. As they have been presented so far, it is pretty clear that the PI is stronger than the NMA: The NMA relies, at bottom, on an intuition, whereas the PI offers concrete counterexamples against realism. So, the PI trumps the NMA. What is more, the PI also seems to offer a direct refutation of the NMA itself: the NMA implies that the predictive success of the false theories is miraculous, but so many miracles cannot be tolerated. Thus the PI seems to refute both realism and the NMA.

I understand the PI as follows. The premise of the PI states that a significant proportion of successful historical theories are false.[8] [9] The conclusion of the PI states that a significant proportion of successful theories, *historical or not*, are false. This inference, which I call "simple PI", has approximately the form of an inference from a subset of a population (past successful theories) to the whole population (all successful theories). Let us then make two simplifying assumptions: First, let us identify approximate truth with truth and count theories that are partly true and partly false, but not approximately true, as false. Second, let us ignore the difference between the probabilities occurring in the formulation of realism and the corresponding relative frequencies. Given these two assumptions, the conclusion of the simple PI – a significant proportion of successful theories are false – is the negation of realism, which states that nearly all successful theories are true. So, the premise of the simple PI supports the negation of realism, – and therefore undermines realism.[10] This is how the simple PI is an argument against realism.

How is the quantifier "a significant proportion" in the premise and the conclusion of the simple PI to be understood, i.e., how many counterexamples of successful-but-false theories does it take to cause trouble for realism and the NMA? Ladyman and Ross (2005) claim that one or two are enough to refute the NMA, because miracles should not occur at all. This strikes me as too strict. The NMA's claim that the predictive success of a false theory would be a miracle is just a vivid way of stating that it is very improbable that a theory be successful, but false. A judgment to the effect that a type of event is very improbable is not necessarily undermined, if one or two events of that type are observed. So the NMA can tolerate one or two exceptions.[11] [12] In line with this understanding of the NMA, realism

---

[7] Many further examples can be found in Timothy Lyons (2002), Kyle Stanford (2006), and Peter Vickers (2013).

[8] In what follows, the locution "a significant proportion" always means "at least a significant proportion". Thus, the premise of the PI includes the possibility that most or all successful historical theories are false.

[9] The premise, in turn, is supported by the sample of past refuted theories offered by the anti-realist.

[10] What I am using here is the equivalence that a statement P undermines a statement R, just in case it supports the negation of R. This equivalence is true on many or most accounts of inductive inference, for example, if undermining and supporting are understood in probabilistic terms: $\Pr(R|P)$ is near zero iff $\Pr(\neg R|P)$ is near one. Sometimes I use the notions of supporting and undermining and similar notions in an absolute sense (in probabilistic terms: probabilities near one or near zero), and sometimes in an incremental sense (in probabilistic terms: increase or decrease of probability). The context will make clear what I mean.

[11] Compare Saatsi/Vickers (2011, p. 33) and Wray (2015).

[12] An analogous remark applies, if the NMA is interpreted as an instance of IBE (inference to the best explanation). IBE is a

should likewise not be interpreted too strictly, i.e., should not be taken to be refuted by one or two counterexamples. However, if the number of counterexamples gets higher than just one or two, then the NMA and realism face pressure fairly quickly. Fortunately, for our purposes we need not attempt to determine the precise number of counterexamples necessary to refute realism and the NMA. Hence, we can avoid the messy business of developing precise criteria for individuating and counting theories (Lewis 2001, Lange 2002).

2.3 Two objections

A common response to the simple PI offered by realists is the following. Our current best theories can generally be taken to enjoy higher degrees of predictive success than past refuted theories, i.e., they can generally be taken to have made better, and more diverse, correct predictions than past refuted theories. Therefore, we can restrict the scope of realism to theories with current levels of success: The inductive inference from the success of a theory to its truth is cogent only when the theory enjoys a level of success typical of our current best theories (or better). This is the form of realism I will defend. Hence, from now on I will understand realism in this restricted sense. Realists may then make two claims. First, because of the difference in degrees of predictive success, past refuted theories no longer provide a reason to think that any of our current best theories are false, in other words (restricted) realism is no longer threatened by any form of PI. Second, the NMA can likewise be restricted to current levels of predictive success. Then it is likewise safe from any form of PI and can be used to support (restricted) realism.

Against this, anti-realists can offer two objections. The first objection, which I call the objection from the past, is directed against the first above claim according to which (restricted) realism is no longer threatened by the PI. Consider a realist in 1900. Such a realist could have reasoned in exactly the same way as the realist of today: “During the history of science scientific theories have generally become more successful. Hence, our current best theories are considerably more successful than past refuted theories. Hence, I need not worry about the theory changes in the past, and can maintain realism for our current best theories.” This reasoning would have been proven wrong, says the anti-realist, because many of the best theories of 1900 were refuted later on. In general, at any point in the history of science, whether 1700, or 1800, or 1900, realists could have reasoned in exactly the same way as the realist does today, namely that their respective best theories have become more successful, and are therefore immune to a PI on the theory changes of the past. But all such reasoning would have been proven wrong by subsequent theory changes. The anti-realist asks: “How does our situation today differ from the situations in the past? Why should current theories be – all at once – safe from theory change?” These questions are rhetorical insinuating that nothing is relevantly different today. Hence, realists are wrong to think that current theories will not be refuted. Instead we should expect that current theories will suffer the same fate as past theories.[13]

Now there obviously *is* a relevant difference between past and present theories, namely a difference in their degrees of predictive success. People in the past could not have reasoned in *exactly* the same way as the realist just did. They could only have reasoned in an *analogous* way, but from lower levels of success. According to the objection from the past, this difference is not decisive. Thus, I understand the objection to suggest an extrapolation: “In the history of science, the levels of success of the best theories have been growing continuously for several centuries now, and theories keep being refuted. Hence, from the refutation of theories that enjoyed past levels of success we should expect that theories that enjoy current levels of success will also be refuted.” This argument is another version of the PI. I call it “projective PI”, because it explicitly extrapolates along degrees of predictive success.[14]

I assume that the projective PI has the conclusion: “A significant proportion of our current best theories are false”. This conclusion undermines realism (in the restricted form), because I assume that,

---

kind of inductive inference and therefore fallible. It is not refuted, if some cases of best explanations turn out to be false, as long as such cases are rare.

[13] This line of thought is presented forcefully by Brad Wray (2013). It is surprisingly complex, so I won’t attempt to analyze all of its aspects here.

[14] Ideas akin to the projective PI are mentioned or hinted at by Stanford (2006, Ch. 1.2), Gerald Doppelt (2007), Bird (2007), Sherrilyn Roush (2009), Wray (2013, p. 4328), and others.

as in the case of unrestricted realism, a few counterexamples (but not one or two) suffice for refutation. Hence, if the projective PI succeeds in establishing its conclusion, realism is refuted. The premise of the projective PI can be taken to be something like: "Up until present levels of success, theories have been refuted at a significant rate".[15] This is quite vague, but it will do for our purposes.

Let us turn to the second objection. The existence of past refuted theories forced the realist to alter his initial position, and he did so by restricting the scope of realism to theories enjoying current (and future) levels of success. The second objection states that this move is ad hoc, made solely for the purpose of saving realism about our current best theories and not supported by any independent reasons.[16] Theories have become more successful throughout the history of science and will probably continue doing so in the future. Why think that it is the precise levels of success enjoyed by our current theories that suffice for truth? Why not think that it is only at some future time, decades or centuries from now, that a level of success sufficient for truth will be reached? Why not give up on realism entirely? As long as the realist is not able to provide an independent reason for restricting the scope of realism in precisely the way he does, namely to theories enjoying current and future, but not past, levels of success, his move has to be judged to be ad hoc.

Realists may reply that they actually *have* an independent reason for restricting realism to current levels of success, namely the NMA. The success of past refuted theories was not a miracle, because their degrees of predictive success were low, whereas the success of our current best theories would be a miracle, if they were false, because their degrees of success has since increased. To this the initial reaction of the anti-realist will be to remind the realist that the NMA is, in fact, no longer available, because it was refuted by the simple PI.[17] But let us assume that the anti-realist allows that the NMA may survive in a modified form. Then she can still object that modifying the NMA in the manner just suggested is ad hoc as well. The NMA, as introduced by the realist, does not talk about degrees of success at all, it only talks about success *simpliciter*. It may be taken to imply that a level of success *exists* that suffices to infer truth, but it does not tell us *how much* success suffices for truth. In particular, it does not tell us that current levels of success suffice for truth. The anti-realist can then maintain that invoking the NMA in the suggested manner only pushes the problem one step back: The realist's claim that the NMA supports realism for current and future, but not for past, levels of success is also ad hoc and not based on any independent reasons.

Given the two objections where do we stand? Realism, as restricted to theories with current levels of success, was meant to be safe from the PI. Moreover, it was supposed to be supported by a restricted version of the NMA. But the objection from the past implies that even this restricted realism is threatened by the incidence of past refutations, namely via the projective PI. Hence, the realist still faces two arguments that have to be balanced against one another, the projective PI and the restricted version of the NMA. Which argument is stronger? On the one hand, the projective PI is less powerful than the simple PI, because it requires extrapolating along degrees of success, and that certainly reduces the strength of the inference. On the other hand, the restricted NMA is considerably weaker than the original NMA, because it is ad hoc. To reach a verdict at this point we would have to try to determine the strength of each argument – certainly not an easy task.[18] Fortunately, we need not attempt to decide the matter, because both objections can be fully rebutted, as I will now show.

---

[15] The projective PI, as I understand it here, has roughly the form of an inductive inference from the set of past cases of a population to the next future cases of the population (similar to the inference from observed ravens to the next unobserved raven, but with the difference that not all past cases have actually been observed, so that there is an initial inductive step from the sample of actually observed past refuted theories offered by the anti-realist to the whole set of past theories). The simple PI has roughly the form of an inductive inference from the set of past cases of a population to the whole population (from observed ravens to all ravens). Note, however, that a number of inferentially relevant features, such as the role of degrees of success in the projective PI and the role of "attrition" (Ruhmkorff 2013), are not captured by the two inductive forms.

[16] Stanford (2006, Ch 1, 6) and Michael Devitt (2011) also mention a charge of adhocness, but they may have something like the first objection in mind.

[17] Thanks to Paul Thorn for the reminder.

[18] At this point one may diagnose a stalemate: Neither the realist nor the anti-realist are able to gain the upper hand (compare Stanford 2006, Anjan Chakravartty 2007). Stanford goes on to develop a new version of the PI, the "New Induction", which he claims avoids the stalemate. The New Induction states that we should project the existence of unconceived alternatives from past to present. The New Induction adds a number of novel and interesting considerations to the realism debate, but some authors argue that it ends in stalemate as well (Kukla 2010, Ruhmkorff 2011, Egg 2016), whereas I think it can be countered with the material presented in this paper, as I will indicate below.

3 The main thesis

3.1 Two premises

My goal will now be to show that the projective PI is unsound. I will do so by showing that the premise of the projective PI ("Up until present levels of success, theories have been refuted at a significant rate") is incompatible with the following claim: *Our current best theories enjoy very high degrees of success, far higher than those enjoyed by past refuted theories, practically all of which enjoyed merely moderate degrees of success at best, before they were refuted*. This is the main thesis of the paper. It will also be used to reply to the second objection of ad-hocness.

My argument for the main thesis has two premises. The first premise states that our current best theories have received big boosts to their empirical success in the recent past and are therefore far more successful today than any theory from the distant past, practically all of which enjoyed moderate success at best. Here the term "recent past" denotes the last few decades, and the term "distant past" denotes earlier times. These definitions are pretty vague, but that will not matter. Some pertinent developments started 70 to 80 years ago, around World War II, hence this is a useful date to have in mind. The argument for the first premise will take up most of the paper.

The first premise does not talk about refuted theories per se. It just compares our current best theories to the best theories of the distant past with respect to their degrees of success. In contrast, the second premise does talk about refuted theories. It states that virtually all refuted theories of the recent past were, before they were refuted, only moderately successful at best. The two premises imply the main thesis of the paper. The first premise implies that practically all refuted theories *of the distant past* were merely moderately successful, whereas the second premise states the same for the refuted theories *of the recent past*. Furthermore, the first premise implies that our current best theories enjoy very high degrees of success.

What do the terms "very high (degrees of) success" and "moderate (degrees of) success" mean? A theory enjoys very high success, if its set of true predictions is very large and highly diverse, or if a number of its predictions agree very precisely with observation, or both. A theory enjoys moderate success, if its set of true predictions is of limited variety, and the predictions and/or the data are of limited precision. If a theory is known to make significant false predictions, it does not enjoy any degree of success. The terms "very high success" and "moderate success" are chosen in order to highlight the contrast between the two classes of theories mentioned in the main thesis, current best theories and past refuted theories. In particular, I take it that many theories considered by scientists to be well-established, or sufficiently successful to be accepted, count as enjoying merely moderate degrees of success. The point of the main thesis is that even among successful theories there are large differences in degrees of success and refuted theories appear only at the lower end of the spectrum.

For the purpose of replying to the two objections talk of "moderate" and "very high" degrees of success is in principle dispensable. First, as we will see, replying to the projective PI merely requires us to assess differences in degrees of success with respect to whether they allow or don't allow the extrapolation of theory failure. Second, replying to the objection from ad-hocness merely requires us distinguish between the success of our current best theories, and the success of past refuted theories in such a way as to make the application of the NMA to the former, but not the latter, plausible. In principle, these two purposes are distinct[19] and call for different kinds of assessment. However the empirical material presented on the following pages will show that talk of "moderate" and "very high" degrees of success is intuitively quite compelling and can serve both purposes.

I will often describe predictive success by using the notion of a test of a theory. As I use the notion, a test of a theory is every occasion in which a prediction of the theory, i.e., a known observable consequence of the theory, is compared with a corresponding observation. If the prediction agrees with the observation, the theory passes the test and receives some measure of success. If the prediction and the observation disagree, the theory fails the test and suffers from an anomaly. If the anomaly is signif-

[19] A further analysis would presumably show that the two kinds of assessments are related somehow, but I discuss them separately in this paper.

icant (or anomalies accumulate), the theory is refuted and a theory change may take place, or so I assume.[20] Given the notion of a test of a theory we can say that a theory enjoys very high success, if it has passed a set of tests that is either, both very large and diverse, or includes very severe tests, or both, and has not failed any significant tests. And a theory enjoys moderate success, if it has merely passed a moderately diverse set of tests of moderate quality, and has not failed any significant tests.

In preparation for the reply to the second objection of ad-hocness I will occasionally use the notion of confirmation. I assume that confirmation is connected with the kind of epistemic probability occurring in the definition of realism: That a theory is absolutely confirmed by some set of observations entails that the theory becomes probable or credible, and that a theory is incrementally confirmed by some set of observations entails that the probability of the theory increases (compare footnote 7). To justify claims that a theory is absolutely (incrementally) confirmed by some observation, we can appeal to the no miracles intuition (in a suitably weakened form). Hence, whenever it is shown on the following pages that some theory becomes more successful we can add an inferential step employing the NMA intuition: Every such increase in the success of a theory implies an increase in its confirmation, and thereby its probability. Hence every use of the notion of confirmation on the next few pages is meant to indicate this inferential step, and is meant to prepare for my reply to the second objection (the thrust of the reply will be that with respect to our current best theories, instances of incremental confirmation add up to very strong absolute confirmation and that this is not the case for past refuted theories). Note that for my reply to the projective PI neither this additional inferential step, nor the notion of confirmation, nor the NMA, is needed.

3.2 Four obstacles

Let us now tackle the first premise, according to which our current best theories have received big boosts to their empirical success in the recent past and are therefore far more successful today than any theories of the distant past practically all of which were only moderately successful at best. My argument for the first premise relies on an observation about the recent history of science. The observation, which will be documented in the next section, is that, generally speaking, there has been an explosion of scientific evidence in the last few decades. The amount and quality of scientific evidence has increased enormously in a large number of scientific areas in this time. Of course, more and better evidence does not automatically lead to theories with very high success. Indeed there are at least four kinds of obstacle that may prevent theories in a scientific area from acquiring very high success, even if scientific evidence is exploding in that area.

Firstly, even if scientific evidence is exploding in some scientific area, scientists may not know any theories that could benefit from all the new evidence and acquire very high success. No interesting theory may exist in that area (there may be just disorder), or interesting theories exist, but scientists may not have found them yet. For example, economics and other social sciences have accumulated huge amounts of data in the recent past, but for large parts of these disciplines scientists have yet to find any very successful theories.

Secondly, even if the amount and quality of scientific evidence in some scientific area is exploding, and scientists know a theory that, in principle, could benefit from all the new evidence, scientists may not be able to recognize that this is so. They could lack the computing power necessary to produce sufficiently many and/or sufficiently precise predictions from the theory. This problem arises frequently for theories from the quantitative sciences that involve equations, e.g., differential equations. Such equations are often hard to solve. If scientists lack the computing power to solve them, they are not able to obtain good predictions from the theory, so they are not able to test it. In such cases, a theory exists that could, in principle, benefit strongly from the explosion of scientific evidence, but scientists are not able to appreciate this.

[20] I assume that these claims about theory testing, as well as similar claims in other parts of the paper, are truisms. They are not accepted by everyone, of course. Many philosophers are still in the grip of well-known qualms of a Kuhnian type. These include the following: "puzzle solving" in normal science cannot confirm the basic assumptions of a paradigm; theories only grow in a "sea of anomalies"; in normal science when a prediction is proven false scientists don't blame the basic assumptions of the paradigm, but rather each other; and so on (see for example Kuhn 1962, Paul Hoyningen-Huene 1993, p. 179, Bird 2000, p. 37). These are interesting assertions to be sure, but I cannot discuss them here.

Thirdly, the explosion of scientific evidence in some area of science may be accompanied by the emergence of many new scientific fields in that area, each field with its own domain and theory, but without any more encompassing theory that covers more than one or a few of the domains of the new fields. Wherever this is the case, many little theories with different domains have to share all the new evidence among one another. Such a pattern can arguably be found in major parts of psychology. In the last few decades psychologists have produced masses of data from numerous psychological experiments of many different kinds. But so far there are no over-arching psychological theories that have benefited strongly from all the data. Rather the data pertain to many different theories of fairly low generality, each with a different subject matter. If that is the case, the success-conferring power of the data is diluted over many different theories, each about a different subject, and despite the masses of new data no individual theory benefits enough to be deemed highly successful.

The fourth obstacle requires a bit more explanation. Scientists typically accept a theory fairly readily, as soon as it enjoys what I called moderate success. They don't wait until the success of the theory is very high. However, once scientists accept a theory, they rarely intentionally test it any more. They consider the rechecking of well-confirmed theories to be a waste of time and energy.[21] Above all, they cannot expect to get recognition for doing so.[22]

Now, after a theory is accepted scientists do normally apply it on many occasions. Many of these applications will constitute tests of the theory. As indicated earlier I understand the notion of a test of a theory in a broad sense: Every occasion in which an observable consequence of a theory is compared with some relevant data, even if only incidentally, is a test of the theory. An accepted theory may come into contact with observation and be tested, if it is used for practical purposes (the practical application may not work), or the theory is used to explain some phenomenon (the theory may not be able to provide an adequate explanation for the phenomenon when it should), and so on. However, it then seems that these kinds of tests are, for the most part, not very severe. They are not especially demanding or careful tests. So even if successful, they do not lead to the theory in question becoming highly successful. It therefore seems impossible for theories to acquire very high success at all, contrary to what the main thesis asserts for our current best theories.

This motivates the following definitions. If a theory is accepted by scientists and used in an application, where the application involves a test of the theory but this is not the main goal of the application, then the testing of the theory will be called "testing en passant". If the test is passed, the resulting kind of confirmation for the theory will be called "confirmation en passant". It is then plausible that, as a general rule, testing of accepted theories is testing en passant, and confirmation of accepted theories is confirmation en passant. Furthermore, let us call a test of a theory which, if passed, leads to a strong increase in success for the theory a "severe test", and a test which, if passed, leads merely to a small or medium increase in success for the theory a "gentle test".

The problem then is this. At the time a theory becomes accepted, it generally enjoys a substantive, but not very high degree of success. Subsequent tests of the theory will normally be cases of testing en passant. They will presumably be gentle tests only, and not capable of increasing the theory's degree of success by a large amount. How then is it possible for a theory to acquire a very high degree of success?

3.3 Our current best theories

Starting in this section I aim to show that there has indeed been an explosion in the amount and quality of scientific evidence in the recent past, and that, despite the four obstacles just mentioned, many theories exist that have been heavily exposed to this new evidence and have experienced very strong boosts to their empirical success. Here are some examples of such theories[23]:

---

[21] Even for results that are not considered to be well-established, scientist are not eager to retest them. "The replication of previously published results has rarely been a high priority for scientists, who tend to regard it as grunt work. Journal editors yawn at replications. Honours and advancement in science go to those who publish new, startling results, not to those who confirm—or disconfirm—old ones." Jerry Adler (2014).

[22] These are empirical claims. I deem them sufficiently plausible to assume them here, but there are exceptions. As we will see later, sometimes scientists do engage in projects where at least one of their aims is to test a well-established theory. Maybe such cases are more wide-spread than it first appears. If so, then all the better.

[23] Remember that realism only asserts the *approximate* truth of these theories. For example, not every infectious disease is

- the Periodic Table of Elements[24]
- the theory of evolution
- the conservation of mass-energy
- the germ theory of infectious diseases
- the kinetic gas theory
- "All organisms on Earth consist of cells."
- $E = mc^2$
- "Stars are giant gaseous spheres."
- "The oceans of the Earth have a large-scale system of rotating currents."
- "There was an ice age 20.000 years ago."
- And so on

It is theories like these that I want to show are far more successful today than any theories of the distant past. They are "our current best theories". Of course, it is entirely infeasible to provide complete descriptions and evaluations of the sets of evidence supporting each of them. Fortunately, that is not necessary. As we will see, partial information about the evidence relevant to these theories is often entirely sufficient to show that they are highly successful. In particular, I will focus on three good-making features of evidence, namely amount, diversity, and precision. I will show that recent evidence provided in many areas of science have far more of each good-making feature than evidence provided historically. For some of our current best theories it is then possible to give a general idea of the whole evidence sets in their support. For most theories I'll discuss, however, I can only provide very incomplete descriptions of these sets. Nevertheless this will often suffice to indicate very high success. Additional support for the big boost in success will come from some more general and indirect clues, e.g., the growth of scientific research in general.

Let me make two remarks about the above list. First, one may object that the second and the last three theories on the list are not, or not mainly, about unobservables.[25] My reply is that my definition of realism is intended to cover such theories as well. The content of many such theories far exceeds the true predictions that are the basis of their success, hence the inference from predictive success to truth is far from trivial, even if the theories are largely or wholly about observables. So, realism about them is far from trivial. What is more, a number of such theories were successful and accepted for some time, but later refuted.[26] Hence, realism about theories concerning observables needs to be defended against the PI as well. The point generalizes. As far as I can see, my list can be taken to include all statements and sets of statements that are covered by the NMA and are threatened by the projective PI with one exception to be noted shortly. Thus, the list not only contains general statements about unobservables, but also statements (and sets of statements) of further quite various sorts, including general statements about observables, and statements about singular events, such as the last entry on the list. I use the term "theory" to refer to all of them.

Second, my list of current best theories does not contain any theories from fundamental physics. Some fundamental physical theories have benefited strongly from the explosion of scientific evidence and are extremely successful today, but there are strong reasons to doubt that they are approximately true. If they are not, they constitute counterexamples to the inference from current levels of success to approximate truth. There are at least two kinds of cases.

First, quantum mechanics and the Special Theory of Relativity have both made numerous correct predictions, sometimes with extremely high precision, but the two theories contradict each other.

---

caused by microbes, viruses or parasites, but the only known exception among more than 200 known types of infectious diseases are prion diseases which are extremely rare. (Thanks to Hasok Chang for this example, even though I use it for a purpose contrary to the one he intended.)

[24] The term "Periodic Table of Elements" is meant to denote the statements that can be taken to be associated with the Periodic Table of Elements, for example the statement that every chemical substance can be decomposed into the chemical elements.

[25] The notion of observability I have in mind here is the usual one of van Fraassen (1980), but the precise understanding of this notion is of little consequence here.

[26] See Stanford (2006, Ch. 2) and Fahrbach (2011a).

Therefore they cannot both be approximately true, or so one might argue.[27] If this is right, at least one of them is a counterexample to the inference from very high success to approximate truth.

Secondly, fundamental physics seems to offer real-life cases of the underdetermination of a theory by all possible evidence (J. Brian Pitts 2011). For example, put crudely, whether gravitation is a field, a force, or a feature of space-time may not make an observable difference. Also, some of the different interpretations of quantum mechanics are observationally indistinguishable (or nearly so).[28] Cases of underdetermination in which the theories involved are highly successful threaten realism: The theories are incompatible, therefore all but one of them are false and constitute counterexamples to the inference from very high success to approximate truth.

Both problems are less threatening than it may seem at first sight, for two reasons: First, both problems are confined to fundamental physics and do not threaten the inference from very high success to truth for theories found in the rest of science, and, second, reasonable modifications of the success-to-truth inference for the domain of theories from fundamental physics are available. Both claims would require extensive argumentation, but I can only offer some brief remarks here.

The first problem is that quantum mechanics and relativity enjoy very high success, but contradict each other. This problem becomes less threatening when the following points are noticed. Both theories are approximately true in quite large domains (although probably not at very fundamental levels). Furthermore, theories from fundamental physics possess certain special properties, profitably analyzed using Bayesian notions. Firstly, such theories are extremely general, i.e., have very large domains. Some aim for maximal generality.[29] In Bayesian terms, high generality makes for small priors. Secondly, theories from fundamental physics often make claims about reality that are highly counterintuitive and conflict strongly with pre-scientific common sense beliefs, e.g., about time, space, causality, simultaneity, determinism and so on. This also makes for small priors. If the priors are very small, it is less surprising that even very high success does not suffice for truth. What we then need is even more empirical evidence than we already possess, and that is, of course, precisely what physicists seek, and what they think will decide at least some of the issues.

My reply to the problem of underdetermination is along standard lines. To begin with, gathering more evidence and seeking to make the theories even more successful obviously doesn't help. There are then two kinds of cases. First, there are the cases in which the rival theories are artificially constructed from a given theory T. For example, given T one may define an empirically equivalent rival T* by stipulating that T* has the same observable consequences as T, but differs about unobservables (e.g., Kukla 2001). Such rivals are entirely contrived, and are not taken seriously by scientists. We can dismiss them as highly implausible by invoking theoretical virtues such as simplicity. This means that the success-to-truth inference is understood as an inference from success-cum-simplicity to truth, where contrived theories count as non-simple. It has then often been remarked that all cases of underdetermination outside of fundamental physics are of the contrived sort (e.g., Ernan McMullin 1984, p. 11, Stanford 2006, Ch 1). Therefore, they can be dealt with as suggested.

The second kind of cases of underdetermination are those mentioned above, which cannot be easily dismissed as contrived, but which are specific to fundamental physics. Here one may note that the competing theories merely differ with respect to relatively few facts, and otherwise share a lot of content. For example, the different interpretations of quantum mechanics agree on the core mathematical structure. One may then argue that it is generally improbable for theories with the same empirical consequences to differ completely on a theoretical level (John Norton 2008). If this is right, the disjunction of the competing theories captures what they have in common and can serve as input for a success-to-truth inference. Beyond that, we may have to live with the fact that some truths about reality are epistemically inaccessible for us humans, possibly including some really interesting ones, e.g.,

[27] Jeffrey Barrett (2003) argues that the claims that quantum mechanics and relativity are approximately true have no clear sense for us today and are "hopelessly vague". See also Barrett (2008).

[28] Claus Beisbart (2009) and Jeremy Butterfield (2013) discuss examples of underdetermination in cosmology. If any of their examples concerns theories with very high success, then they also threaten the success-to-truth inference. I treat cosmology as part of fundamental physics.

[29] Fundamental physics is the only scientific discipline (apart from metaphysics) that aims for maximal generality, at least in principle. Besides, it is the only scientific discipline that does not rest content with approximate truth, but aims, at least in principle, to find theories that are completely precise and exception-less (compare Earman/Roberts 1999, 446), which is another reason to assign very low priors.

concerning the truth of determinism (on which different interpretations of quantum mechanics disagree). But again, this sort of inaccessibility is confined to fundamental physics, and absent from other areas of science.

Much more could be said on this topic, but what has been said suffices to show that the problems for realism in fundamental physics are special to this area. They don't affect realism in other areas of science, and can be safely ignored when discussing realism there. So, theories from fundamental physics don't occur on the above list of current best theories and are set aside in this paper.

4 The explosion of scientific evidence

4.1 The exponential growth of scientific research

An important phenomenon that helps to explain the recent explosion in scientific evidence is the growth of the total amount of scientific research over the history of science. Let us take a look at this growth. It is plausible to assume that the total amount of scientific research done by all scientists in some period of time is very roughly proportional to two other quantities, the total number of scientific journal articles published by all scientists in that period, and the total number of scientists active in that period. What we then observe is that in the last 300 to 400 years the total number of scientific journal articles per year has, in general, been growing in an exponential fashion, with a doubling rate of 15 to 20 years. Roughly the same growth rate can be observed for the total number of scientists (although in this case the data record is a bit sketchy for times before the 20th century).[30] Hence, we can conclude that the over-all amount of scientific research done by all scientists per year has also increased with a doubling rate of 15 to 20 years (Figure 1). This is a very significant rate of growth. It means, for example, that around 90% of all scientific research ever done in the whole history of science has been done after World War II.[31] This suggests that in most scientific fields most of the growth in scientific evidence occurred in the recent past.

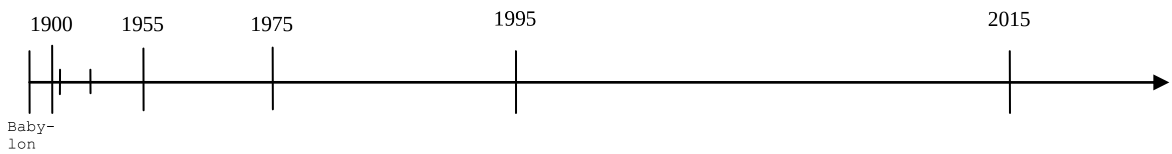

Figure 1. A depiction of the time-line in which the length of any interval is proportional to the amount of scientific research done in that interval. The depiction assumes a doubling rate of 20 years.

4.2 Big data

In this subsection I show that the *amount* of many kinds of data has grown very strongly in the last few decades. We can distinguish two sorts of cases. First, in some scientific disciplines scientists themselves collect most of the data. For instance, synthesizing new chemical substances is still done mostly by chemists, and searching for fossils is still done by palaeontologists. For such disciplines it is plausible that the amount of data has grown very roughly proportionally to the number of scientists. If we assume a doubling time of 20 years for the number of scientists (which is a conservative estimate), then there are around 30 times as many scientists in 2000 as in 1900, around 1,000 times as many as in 1800, and around 30,000 times as many as in 1700.[32] It is therefore plausible that in disciplines in which the scientists themselves collect the data, the amount of data has often very roughly increased in a similar fashion.

[30] De Solla Price (1963), Vickery (2000), Fahrbach (2011a).

[31] A growth with a constant doubling rate can be represented by an exponential function $f(t) = a\, e^{bt}$, where t is time and a and b are suitable constants. A doubling rate of 20 years means that, approximately, $f(t+70 \text{ years}) = 10 \cdot f(t)$ and $f(t+100 \text{ years}) = 30 \cdot f(t)$.

[32] See previous footnote.

A good example is the growth of the fossil record. Take dinosaur fossils. Wang and Dodson (2006) observe that from 1990 until 2005 the number of known genera increased from 285 to 527. "Since 1990, an average of 14.8 genera have been described annually, compared with 5.8 genera annually between 1970 and 1989 and 1.1 genera annually between 1824 and 1969." (p. 13601). In the last ten years, the number of genera has grown to 1272 (Starrfelt and Liow 2015). This sort of growth is typical for the growth of the fossil record in general.[33] One hundred years ago, the number of known fossil genera was in the thousands, in the 1990s it was in the hundreds of thousands (Richard Leakey and Roger Lewin 1996, p. 45). Later on I discuss another example, the increase in the number of synthesized chemical substances.

Secondly, in the very recent past data generation has increased much faster than the growth of scientific manpower, because data collection has been *automated*. ”Today you can put instruments practically anywhere. Vast numbers of sensors monitor an equally vast range of phenomena, on every scale, from elementary particles to individual birds to Antarctic ozone levels to the solar wind. These sensors pour colossal volumes of digitized data into disk drives" (Paul Edwards 2010, p. xix). The automatic gathering of data has led to a huge increase in the amounts of data in many scientific fields. Thus, the total amount of scientific data is estimated to double every year today (e.g., Szalay and Gray, 2006, p. 413-4).

Here are three examples, from astronomy, genetics, and oceanography. Since 2000, the Sloan Digital Sky Survey has automatically scanned one-third of the entire sky, and thereby measured the precise brightnesses and positions of hundreds of millions of galaxies, stars and quasars. By contrast, one of the first big surveys of the sky using photo plates, conducted at Harvard and completed in 1908, measured the brightnesses and positions of 45,000 stars. Another example is the data produced by DNA sequencing. The most important database for registering decoded sequences is GenBank. Since its inception in 1982 it has grown with a doubling rate of 18 months and harbours around 190 million sequences today.[34] A final example is the Argo system. Installed in the years 2005 to 2009 the Argo system consists of a network of 3,600 robotic probes that float in the Earth's oceans. The probes continuously measure and record the temperature, salinity and velocity of the upper 2,000m of the ocean, surfacing once every 10 days to transmit the collected data via satellite to stations on land. These kinds of growth are only possible with the heavy use of, and an ever increasing role for, automation.

The first obstacle to higher degrees of predictive success of theories is that even if the amount of data has exploded in a scientific area, scientists may not know any theories that could have benefited from it. We can now address this concern, at least for the examples just presented (data about fossils, stars, ocean water, and DNA). In each of these examples the data provided supports a specific theory on my list. So far, so good. However, one might object that the mere amount of data relevant to a theory can, by itself, only be taken to be weakly indicative of its degree of predictive success or confirmation. Therefore I will shortly examine two further good-making features of data sets that are more directly connected with the degree of success of a theory: their diversity and precision.

4.3 Computing power

Let us now turn to the second obstacle facing the very high success of theories: Even if scientists have gathered lots of data in some scientific area, and know a theory in that area for which the data are, in principle, relevant, they may not be able to bring the theory and data in contact to a sufficient extent, because they are not able to derive sufficiently many and/or sufficiently precise predictions from the theory. In most physical sciences, and increasingly in many other sciences, theories are formulated in quantitative form using mathematical equations, e.g., differential equations. In order to obtain predictions from such theories, the equations have to be solved, and calculations such as determining the values of functions, integrals, and so on have to be made. These mathematical and computational tasks are often very demanding. To what extent they could be accomplished at a given time in the history of science is a question of the computing power of scientists at the time, i.e., the hardware and software available, where software includes all methods and algorithms for solving equations and making calculations.

[33] On tyrannosaur fossils see Brusatte et al (2010, p. 1481), on human fossils see Bryson and Roberts (2003, Ch. 28).

[34] https://en.wikipedia.org/wiki/GenBank. Compare James R. Lupski (2010).

So, let us examine how computing power has been growing over the last few decades. Firstly, as is well known, the hardware power of computers has doubled roughly every two years in the last 50 years.[35] This is an extremely high rate of growth, much higher than the growth of scientific manpower and scientific publications. Secondly, software progress, although not as easy to quantify, has often been just as dramatic. Computer science, mathematics, and other disciplines engaged in devising computational methods and algorithms have, like the rest of science, grown exponentially. As a result, for many kinds of computational problems the number and efficiency of methods and algorithms has much improved in the last few decades.[36]

The advances in computing power have naturally had an enormous impact on science in a number of ways. They have led to huge improvements in the efficiency of many practical tasks of scientific research such as communication between scientists and the publishing of scientific results.[37] In particular, they have helped to make the practical tasks involved in the testing and confirmation of theories much easier. They have completely transformed how scientists gather, store, manage, and search data. This is witnessed by our examples of automated data gathering in astronomy and ocean science. Finally, concerning the second obstacle, the improvements in computing power have had a huge effect on the ability of scientists to derive predictions from quantitative theories. Generally speaking, scientists are now able to produce a much higher number of more diverse, more accurate, predictions from such theories, with much less effort than was required previously. These advances are so evident that I can omit examples. The upshot is that in a large number of scientific fields the second obstacle on the road to very high degrees of predicative success has been completely removed.[38]

An important consequence of the enormous improvement of computing power is that rechecking quantitative theories has often become extremely easy. This is relevant for the fourth obstacle: The problem that once scientists firmly accept a theory they rarely put much effort into further testing it. This seems to preclude further tests from being severe, i.e., from leading, if passed, to much higher success. However, because of the huge increase in computing power, there are numerous fields in which scientists are now able to derive predictions, even very precise predictions, from quantitative theories with barely any effort at all. Hence, if relevant data are available, the testing of theories, even the severe testing of theories, has often become extremely easy. This is illustrated by a remark in a textbook on computational fluid mechanics: “It is now possible to assign a homework problem in computational fluid dynamics, the solution of which would have represented a major breakthrough or could have formed the basis of a Ph.D. dissertation in the 1950s or 1960s.” (Tannehill et al. 1997, p. 5, cited in Humphreys 2004, p. 49). In the quantitative sciences this kind of progress is entirely typical.

4.4 Diversity of data

The third obstacle to very high degrees of predictive success is when the explosion of scientific evidence in some area merely leads to an increase in the number of theories with different subject matters in that area, so that many little theories with different subject matters have to share the confirming power of all the new evidence. I will consider two ways to overcome this obstacle, diversity, and precision of evidence. In this sub-section, and the next, I discuss diversity of evidence. In section 4.6 I discuss evidential precision.

The point of diversity of evidence is that if many different pieces of evidence are relevant for one and the same over-arching theory, then the confirming power of the evidence converges on that theory instead of being diluted over many different theories with different domains (cf. Markus Eronen 2015). A theory that is unifying in this sense is able to become enormously successful.

Consider the example data sets presented above. Clearly each of them is considerably diverse. For example, astronomical projects such as the Sloan Digital Sky Survey don’t record features of the same star again and again, but record features of hundreds of millions of *different* stars and galaxies. Like-

[35] See, for example, William D. Nordhaus (2002).

[36] For some anecdotal evidence see Robert Bixby (2002, p. 14), Holdren et al (2010, p. 71) and Eliezer Yudkowsky (2007).

[37] For example, the data of ARGO, GenBank, the Sloan Digital Sky Survey, and many other surveys are freely accessible online for everybody.

[38] Humphreys (2004, especially Chapter 3) discusses the huge difference the increase in computing power has made in the physical sciences. He observes that “much of the success of the modern physical sciences is due to calculation” (2004, p. 55).

wise, the fossils gathered by palaeontologists come from a wide variety of different locations, strata, and species, and the probes of ARGO are distributed evenly over the oceans of the earth.

But now a worry arises. Although each of these data sets exhibit *some* diversity, the data of each set are nonetheless all of the same, even if broad, *kind*. When a data set is homogenous in this way, its diversity is limited. For example, the diversity of the data from the Sloan Digital Sky Survey is limited because it relies entirely on visible light, and the diversity of the data from fossils is limited as they are all rocks. One may then well wonder how such data sets can ever provide a theory with very high success.

To describe this worry let us introduce a rough distinction between two sorts of diversity, regular diversity and deep diversity. A data set is *regularly diverse*, if all its data are of the same, more or less broad, kind. For instance, all the data may result from the same kind of causal process, e.g., the detection of visible light from stars, or may originate from the same kind of phenomena or objects, e.g., fossils. Such data typically vary with respect to certain parameters. For example, fossils differ with respect to form, location and strata, and DNA sequences stem from different organisms, species and loci in the genome. A data set is *deeply diverse*, if it is the union of disjoint subsets where the data of different subsets are of fully different kinds (in that case the subsets are usually regularly diverse.) The worry then is that the data sets above are only regularly diverse, not deeply diverse, and regular diversity does not suffice to provide theories with very high success.

The worry can be dispelled. First, exhibiting only regular diversity does not necessarily preclude a data set from offering very strong confirmation for a theory. The amount of confirmation depends on the specifics of the particular case. If, for example, a theory covers phenomena that can be measured very reliably, and the theory does not go too far beyond those phenomena, then regular diversity can suffice to provide very strong confirmation. For example, the measurements of the ARGO-system, although being merely regularly diverse, quite clearly strongly confirm the theory that the oceans of the Earth contain large-scale systems of rotating currents.[39]

Second, for most of our current best theories, the relevant data sets are actually deeply diverse. A good example is the theory of evolution. Data from fossils and DNA sequences are just two kinds of evidence supporting the theory. Further support comes from many very different kinds of evidence from biogeography, comparative anatomy, embryology, genetics, different divisions of molecular biology, and so on. That deep diversity of evidence is decisive for the confirmation of the theory of evolution has been pointed out many times. We have to be grateful to creationists here who forced biologists to make the evidence for evolution more explicit than is common in science for highly confirmed theories.[40]

Furthermore, in the last few decades a precipitous increase in the diversity of instruments and measurement techniques has led to a strong increase in the deep diversity of evidence. Here are two telling examples. Before the 1930s scientists had only light microscopes. Subsequently they have developed many further kinds of microscopes such as electron microscopes, scanning probe microscopes, acoustic microscopes, X-ray, infrared and ultraviolet microscopes, and so on. Likewise in astronomy: Before the 1930s astronomers had only light telescopes. Subsequently they have developed many different kinds of instruments, which not only cover most of the electromagnetic spectrum, from radio-waves to gamma-rays, but also detect neutrinos, muons, Oh-My-God particles, and other things.[41] Hence, for many of our current best theories, the relevant data sets exhibit diversity of the deep kind.[42]

At this point I can offer a further response to the problem of testing en passant, i.e., the problem that once scientists have accepted a theory they usually stop intentionally devising severe tests of it. In

[39] A reminder: Judgments about the degrees of success of theories depend only on the amount, quality, and diversity of the passed tests, and are independent of the NMA intuition, whereas judgments about the confirmation of theories by evidence concern the probability or credibility of the theories given the evidence, and depend on the assumption that the NMA has probative force.

[40] See any textbook on the theory of evolution, sites like www.talkorigins.org, or the entry "Evidence of common descent" in Wikipedia.

[41] The Sloan Digital Sky Survey mentioned earlier is just one of dozens of sky surveys produced in the last decades, see Djorgovski et al (2012).

[42] To get an idea just how great the diversity of currently available measurement devices and techniques is have a look at Wikipedia, e.g. the entries on "measuring instruments" and "non-destructive testing".

all examples offered so far (fossils, ARGO probes, further types of telescopes and microscopes, and so on), practically all data have been collected *after* the respective theories were accepted. Hence, the amount and diversity of evidence in support of these theories has increased enormously, even though they were already accepted and, for the most part, no longer intentionally tested. Most of the tests corresponding to these data do not constitute severe tests of the respective theories. But the whole sets of data often exhibit deep diversity or very strong regular diversity, and are therefore able to strongly confirm their respective theories.

4.5 Core theories

Let us consider diversity of data from a different angle. Let us consider theories that are highly unifying in the sense that they are involved in much of what is going on in their respective scientific discipline and possibly in further disciplines. I will call such a theory a "core theory" ("CT") of its discipline, and every situation in which a core theory is involved somehow an "application" of the core theory. For example, the theory of evolution plays a role in a large number of situations in most fields of biology. As Theodosius Dobzhansky famously said: "Nothing in biology makes sense except in the light of evolution" (1973, p. 125). Another example is the conservation of mass-energy which has a huge number of applications in physics, chemistry, engineering, and so on.

Many applications of CTs don't constitute tests of the respective CT. However, in the case of our current best CTs a large number of their applications typically do constitute tests, i.e., they involve predictions to which the CT contributes and which come into contact with observation (examples in a moment). Given my broad understanding of the notion of a test, such applications count as tests of the CT. Most of these tests will merely be gentle tests, i.e., will impart, if passed, merely small or medium increases in success on the CT. One reason is that many of our current best CTs have been accepted for quite some time, hence the testing is mostly testing en passant. An important kind of case is this: A large amount of research in many scientific disciplines is concerned with adding details to the respective CTs. In such cases the CT will typically be part of scientists' theoretical background. As such it will typically imply constraints about the details scientists are seeking. It will lead scientists to form expectations about the properties of the details. Mostly these expectations will not be very specific. For instance, based on the theory of evolution, biologists have expectations of various sorts about the features of the objects they study such as organisms, fossils, genomes, etc., and these expectations are mostly not of a very specific sort. Hence, the tests are merely gentle, and if the expectations are met, the gains in success are merely small or medium.

However, we can now once again use the fact that in practically all scientific disciplines the amount of scientific research has, by and large, been growing exponentially. As a consequence the number of applications of CTs has increased enormously. What is more, the *variety* of applications has risen just as strongly, because the increase in scientific research has been accompanied by an unremitting *diversification*. All the time we see scientific fields splitting into sub-fields and new scientific fields emerging. This has led to an enormous rise in the number of scientific fields in most scientific disciplines (Wray 2011, 117). Hence, current CTs typically play a role in a large number of scientific fields, each with its own domain, questions, and methods. In contrast, although some of the theories on Laudan's list were also CTs of their respective scientific discipline, they were involved in far fewer fields than current CTs, simply because far fewer fields existed at those times. So, current CTs have a much higher variety of applications than past CTs. In the case of our current *best* CTs a large portion of these applications have been tests. These tests have mostly been passed. Many or most of these tests were presumably merely gentle tests, in which the theories could only become more successful by a small or medium amount. But because their number and diversity has been so large, they have, taken together, massively boosted the success of our current best CTs

The foregoing considerations are of a fairly general nature and need to be qualified a bit. Strong growth of the amount of scientific research in a scientific discipline and concomitant growth of the diversity of scientific fields in that discipline do not automatically result in much higher success for the CTs of that discipline. Both kinds of growth have been quite universal phenomena occurring in all of science. CTs may exist for which the large increase of both scientific work and diversity of fields has not led to much more success. There may have been no explosion of scientific data in the area, or there may have been one, but the new data may not be of the right sort. Another kind of case is rational choice theory, which is a CT of economics for it has been applied to a large number of economic problems, yet most of these applications arguably don't constitute tests of it. Nevertheless, despite these qualifications, the material of the last few pages strongly supports that many CTs do exist that have profited hugely from the growth of scientific research and the ever-increasing diversification. Let me offer two further examples.

The first example is plate tectonics, a CT of geology. Firmly accepted by geologists for five decades now, it has a large number of applications in many fields of geology: It is involved in explanations of the volatility of volcanoes, the shapes of mountains and sea-floors, the magnetic stripes on the sea-floor, the distribution of fossils, the distribution, frequency and approximate strength of earthquakes and tsunamis, and so on.[43] As a result, it has passed a large breadth of mainly gentle tests that have provided it with very strong confirmation.

The second example is the CT of chemistry, the Periodic Table of Elements. The Periodic Table has been fully accepted by chemists for many decades now. Testing it has not been on their minds in the recent past, hence we are dealing with testing en passant here. However, the Periodic Table is highly unifying in the sense that it plays a role in practically everything chemists do and think. It provides the basis for the systematic categorization of all chemical substances in terms of their molecular structure, and it implies constraints about the observable features of every chemical substance and every chemical reaction. Hence, each newly synthesized chemical substance and each new chemical reaction amounts to a gentle test of the Periodic Table.[44]

Now as in the rest of science, manpower in chemistry has risen exponentially. Today, millions of chemists work in tens of thousands of fields. Accordingly during the past 200 years the number of newly synthesized chemical substances has risen exponentially with a surprisingly stable doubling time of about 13 years (Joachim Schummer 1997). The Chemical Abstracts Service (CAS), which catalogues all known chemical substances (except macromolecules such as DNA sequences and proteins) is now registering around 10 million new chemical substances every year. By 2015 it has recorded 100 million substances.[45] The huge number of newly produced chemical substances and reactions means that there have been a huge number and diversity of gentle tests of the Periodic Table. All these gentle tests have been passed, as witnessed by the fact that the Periodic Table has been entirely stable for more than 100 years. Each test may only have led to a weak or medium increase in success, but together they have led to a huge over-all increase in success.[46]

4.6 Precision of data

The third obstacle that may prevent the explosion of scientific evidence in some area of science from leading to enormously successful theories in that area consists in the confirming power of all the new evidence being diluted over many different theories with different domains. A second way in which this obstacle may be overcome, in addition to diversity of evidence for the same unifying theory, is when scientists succeed in producing more precise data from more precise measurements. In that case the problem of dilution obviously does not arise, because if data about a certain object or phenomenon are relevant for a theory, then more precise data about the same object or phenomenon are relevant for the theory as well.[47]

Data become more precise when scientists improve upon existing, or develop new, kinds of instruments and measurement techniques.[48] Enormous progress in both has occurred in the last few dec-

---

[43] Some of these areas are quite different from others, therefore the diversity of the evidence for plate tectonics is of the deep kind.

[44] Further kinds of gentle tests in chemistry are described by William Goodwin (2012, pp. 436/9, 441, 2013, pp. 808/9).

[45] https://www.cas.org/news/media-releases/100-millionth-substance. See the diagram there.

[46] Here is a project for a truly ambitious anti-realist: Conceive an alternative to the Periodic Table of Elements that is entirely different from it, but also able to provide a systematic categorization, like the categorization by molecular structure, of the 100 million chemical substances synthesized so far.

[47] Judgments about the precision of measurement are theory-laden to some extent (and likewise judgments about the size or diversity of data sets). I don't think that this is a serious problem, but cannot argue the case here. Let me just note that the accuracy of scientific data is usually not taken to be threatened by a pessimistic induction in the way truth about scientific theories is.

[48] Devitt (1997, 2008) argues against the PI by invoking the improvement of scientific methods over the history of science, in which he means to include the improvement of instruments and measurement techniques. See also Roush (2009). However, note that replying to the PI by merely pointing out that scientific methods have been constantly improving runs into an objec-

ades. In many scientific areas instruments and measurement techniques have improved enormously, either by big leaps, or by numerous smaller steps, or both. Hence, many kinds of data are far more precise than they were some decades ago. Examples abound. Let me present two brief ones and three long ones. First, the resolution of electron microscopes has improved by several orders of magnitude since their inception in the 1930s. Second, the Hubble telescope is more than 10 times more precise than the best earth-bound telescopes (which are limited by the effects of atmospheric turbulence).[49]

As a third example consider again plate tectonics. An important aim of geology is to determine details of tectonic plates such as their precise boundaries, the velocity and direction of their movements, and so on. For this purpose geologists need precise measurements of the distances between places on the surface of the earth, and the temporal development of these distances. The measuring techniques available prior to the 1980s typically required years to produce meaningful data. In the 1980s, the development of GPS and similar techniques has made distance measurement much more precise. As a result, determining how tectonic plates currently move has become very easy and very reliable. What is more, the thousands of GPS stations dedicated to this purpose are distributed all over the surface of the earth, hence the whole data set exhibits a large diversity[50].

The example shows once again that confirmation en passant can be very strong confirmation. Geologists mostly accepted plate tectonics already in the late 1960s. Hence, the main purpose of the GPS measurements has not been to reconfirm plate tectonics, but rather to find out details about the tectonic plates. Nevertheless, because the data are much more precise (and more diverse) than previous data, they strongly reconfirm plate tectonics. The example generalizes. Very strong increases in the precision of measurements have often resulted in very strong confirmation of theories, even if the theories were already accepted. Thus, despite what we suspected earlier, testing en passant is not necessarily gentle, but is often very severe and often produces big boosts to the success of theories.

The fourth example concerns Einstein's equation $E = mc^2$. In a test in the year 2005 the nuclear binding energy of a neutron in the nucleus of the silicon isotope ^{29}Si was measured in two ways: One team (in the U.S.) determined the atomic-mass difference between ^{29}Si and (^{28}Si + free neutron). The other team (in France) determined the energies of the γ-rays emitted when ^{28}Si nuclei capture a neutron. A second pair of tests did the same thing for two adjacent isotopes of sulphur. These tests reconfirmed $E = mc^2$ with an accuracy of at least 0.00004%.[51] This accuracy is 55 times higher than the accuracy of the previous best test from 1991, which used an entirely different method, namely comparing the sum of the electron and positron masses to the energy released in the annihilation of electron/positron pairs.

This example shows that even in the case of our current best theories scientists do not always stop devising intentional tests. As remarked several times, the rechecking of well-confirmed theories is mostly a side-effect of scientific activity. However the last example shows that scientific projects do exist in which the main purpose, or one of the main purposes, of the project is to recheck a well-established theory.[52] Such intentional tests, if passed, are not considered to be great advances. They are rarely in the limelight of the scientific community, precisely because the theories are already considered to be well confirmed, but they are still sometimes performed when the occasion arises. And it does time and again, because of the development of new instruments and techniques, and the results are noticed and acknowledged. It is simply interesting, and also satisfying, to see that a theory you already accept passes further more stringent tests. And it is always possible in principle that the theory

---

tion analogous to the projective PI that theory change should be extrapolated along the improvement of method.

49 A consequence of the massive improvements of instruments over the last few decades is that the replication of experiments has oftentimes become extremely easy. This means that the corresponding rechecking of theories has oftentimes become extremely easy. For example, most experiments at the cutting edge of physics before WWII can now be performed by students in their university education. (Compare the remarks at the end of the sub-section on computing power.)

50 The diversity is of the "regular" sort, because the data are all of the same kind. They vary with respect to the parameter of location.

51 Simon Rainville et al (2005). See Kim Krieger (2006) for the story behind this validation of $E = mc^2$ in which the two teams made their respective measurements entirely independently from each other until they deemed them stable, and then simultaneously faxed the results to each other.

52 Another example of an intentional test of a well-confirmed theory is a recent test of Newton's law of gravitation for masses separated by 55μm (Clive Speake 2007). This test also served to rule out some versions of string theory, hence testing Newton's law of gravitation was not the only purpose of this project.

fails the test, in which case, if you are really able to show this, your career will prosper mightily. Of course this does not happen. Still, every such passed test results in additional success for the respective theory.

The last example is more general and provides a more indirect clue for the increase in success of our best theories: the increase in precision of time measurement. Since the 1950s the precision of atomic clocks has increased by around one decimal place per decade (Sullivan, 2001, p. 6). Such a growth is also exponential. Thus, in the 1950s, the best clocks reached a precision of 10^{-10}, or one false second in 300 years, and the best clocks today, so-called optical clocks, reach a precision of 10^{-18}, or one false second per 10 billion years (Ushijima et al 2015). Precise measurements of time are vital for numerous scientific purposes, including GPS for example. They have plausibly contributed strongly to the increased success of many theories.

5 Saving Realism

5.1 Completing the argument for the main thesis

Let me recapitulate the argumentation so far. I have shown that in the last few decades the amount, quality, and diversity, of scientific evidence have increased dramatically in a large number of scientific areas. Four kinds of obstacles may, in principle, preclude the explosion of scientific evidence in a scientific area from producing theories with very high degrees of success. However I showed that there are many theories in many scientific areas to which the four obstacles fail to apply. As such, they have benefited hugely from the explosion of scientific evidence. In the last few decades these theories, our current best theories, have passed an enormous number of very severe tests of great variety, and are therefore extremely successful today.

In contrast, in the distant scientific past, i.e., more than a few decades ago, the amount, quality and diversity of evidence was far lower. Generally speaking, the data sets of those times were comparatively small, less diverse, and less precise. And due to the very limited computing power available to scientists at those times only comparatively few and imprecise predictions from theories could be obtained. Therefore, the best theories of those times were merely subject to a moderate number of moderately diverse tests, tests that were moderately severe at best. It follows that practically all theories of those times, whether refuted later on or not, enjoyed only moderate levels of success at best. In sum, we have established the statement that our current best theories have received big boosts to their empirical success in the last few decades and are therefore far more successful today than any theories of the distant past practically all of which were only moderately successful at best. This statement is the first premise of the argument for the main thesis.

The main thesis asserts that our current best theories enjoy very high degrees of success, whereas practically all *refuted* theories were moderately successful at best. Therefore let us now tend to the refuted theories. As we just saw, all theories of the distant past were only moderately successful at best. Hence, this also holds for all *refuted* theories of the distant past. Importantly, this includes almost all examples of theory refutation discussed in the philosophical literature, because almost all of these occurred before the last few decades. For example, all refutations of theories on Laudan's list occurred before 1900.[53]

The question then remains whether there are any theories in the *recent* past that were accepted for some time and benefited from the explosion of scientific evidence, but were subsequently refuted. Let us call such theories "RATs" for "recently abandoned theories". Thus, RATs are defined to be theories of the recent past that enjoyed more than moderate levels of success before they were refuted. Then the second premise of the argument for the main thesis states that there have been virtually no RATs. This premise receives considerable support from the fact just noted that almost all cases of refutations discussed in the philosophical literature occurred in the distant past: It is plausible that the philosophical literature discusses those theory changes in the history of science that are the most interesting and salient and therefore potentially most worrisome for the realist. That practically all of these occurred in the distant past supports the claim that there are no RATs. Furthermore I have kept an eye out for theo-

[53] See Park (2011) and Fahrbach (2011a, 2011b).

ries that were refuted in the last few decades. Unsurprisingly there have been quite a number of refutations, especially at the "research frontier", but I found none that were convincing examples of RATs, i.e., none that enjoyed more than moderate success. Hence, I feel confident that none (or next to none) exists.

Some anti-realists may not be convinced yet. They may search for RATs themselves. Let me kindly make some suggestions for their search. It would not help much to survey articles from scientific journals such as SCIENCE and NATURE, because research articles of this sort normally report findings from the research frontier. At the time of their publication such findings typically enjoy what I called moderate, rather than very high, degrees of success (even if scientists often regard them as sufficiently confirmed to accept them right away).[54] Instead the anti-realist may inspect scientific textbooks. Suitable candidates for examination might be textbooks from disciplines such as chemistry, biology, or astronomy published in the 1960s, 1970s, and 1980s, say. For example, Butterfield (2013) reports that he compared seven astronomy textbooks written over the last four decades with respect to their descriptions of the over-all thermal history of the universe (first dense and hot, then expanding and cooling down, and so on). Unfortunately for the anti-realist, the descriptions largely agree with each other.[55] Butterfield reports a similar agreement for theories of stellar structure and stellar evolution.

It would be particularly advantageous to the anti-realist, if she could present theories that were not only considered to be highly confirmed textbook knowledge at their time, but were *core theories* of their respective disciplines and were later refuted. For instance, an overthrow in recent biology comparable in depth to the change in the 19th century from creationism to the theory of evolution, or changes in recent planetary astronomy like the change from the geocentric to the heliocentric system would certainly help to further the anti-realist's cause. Finally note that the anti-realist has to present more than one or two RATs, because, as pointed out in Sections 2.2 and 2.3, one or two counterexamples do not suffice to defeat realism. As long as anti-realists cannot present a significant number of RATs, I can hold on to the second premise that there have been virtually none of them.

Put together, there are practically no refuted theories, from any period of science, that enjoyed more than moderate levels of success. This establishes, in concert with the first premise, the main thesis of the paper: all our current best theories enjoy very high levels of success, whereas all refuted theories were moderately successful at best.

5.2 Rebutting the projective PI

We can now rebut the projective PI. The premise of the projective PI states that up until present levels of success theories have been refuted at a significant rate. This premise is now proven wrong. The main thesis shows that there is a large difference between the degrees of success of the refuted theories and the degrees of success of our current best theories.

What happens if we provide the projective PI with a premise that is compatible with the main thesis, for instance: "There were many moderately successful theories that were subsequently refuted"? We can assume that this new premise is true. However, it obviously does not support the conclusion of the projective PI that at least a significant proportion of our current best theories are false. The merely moderate success of refuted theories means that they passed only a moderate number of moderate quality tests. That many such theories subsequently turned out to be false does not support the inference to the falsity of a significant proportion of our current best theories, which have each passed a very large and diverse set of, often much more stringent, tests. An extrapolation of theory failure over such a big difference in empirical success is simply not plausible.

The new premise is true, but it is deficient, because it omits relevant information, namely information about the fate of the best theories in last few decades. If we take this information into account, the projective PI becomes even less plausible and an optimistic induction looks much more attractive. Such an optimistic induction has the potential to be a powerful argument in support of realism, but it

[54] Vickers (2013) offers some examples of refuted or abandoned theories from the recent past: S-matrix theory, Steady state cosmology, Velikovsky, Minkowski's theory of the momentum of light. None of these theories were highly successful, they were not even generally accepted or constituted textbook knowledge, hence none were RATs. An especially convincing local version of PI which concerns results of medical studies (hence not RATs) is presented in Ruhmkorff (2013).

[55] Hence we need not decide whether or not these theories actually enjoyed very high degrees of success.

requires careful elaboration which I cannot provide here, hence I will only briefly sketch it and not use it outside this sub-section.[56]

We can describe the additional information with the help of a very simple model. So far we distinguished between two categories of success, moderate success and very high success. It is then plausible to assume that in between these two categories there is a third category of success containing theories with high, but not very high degrees of success. Let us abbreviate the three categories with MOD, HI, and VHI (for "moderate", "high", and "very high"). Every theory that is successful at some time falls into one of the three categories at that time. The idea then is that our current best theories were mostly elements of MOD when they were first accepted, then moved through the intermediate category HI during the explosion of evidence that greatly boosted their degree of empirical success, and are elements of category VHI today.[57]

The main thesis tells us that every successful, but refuted, theory was an element of category MOD before it was refuted, and none made it into categories HI and VHI. Hence, we have a clear pattern: Refutations occurred in category MOD, but stopped occurring for theories that became more successful and moved into categories HI and VHI. This observation supports the claim (and this is the optimistic induction) that our current best theories, i.e., the current members of category VHI, are empirically adequate, i.e., have no false empirical consequences. From this we can infer that these theories are true, for otherwise, if any of them were false despite their empirical adequacy, they would constitute, together with the corresponding true theory, cases of underdetermination of theory by all possible evidence. But this contradicts the widely held view, which I follow here, that outside of fundamental physics cases of underdetermination of this sort don't occur. So, our current best theories are true. It follows that the conclusion of the projective PI, that at least a significant proportion of our current best theories are false, is undermined by the history of science. Thus, if we take into account, as we should, all information about the past, including information about the recent past, any projection of theory failure from past to present levels of success becomes entirely implausible.[58] This completes my argument against the projective PI.

To illustrate my arguments against the projective PI think of people like Duhem, Poincare, and Boltzmann at the end of the 19th century who worried because of cases of theory change in their past. Assume we had asked them in 1900 to imagine the kind of explosion of evidence and the kind of growth in success of the best theories that has actually occurred between 1900 and today. Assume we had asked them to judge whether the occurrence of theory failure at their times can be extrapolated to the very high levels of success of the best theories today. Surely they would have agreed that such an extrapolation was not warranted. Now assume that we had asked them additionally to imagine that practically none of best theories in the decades before the present were refuted. Surely they would have agreed that, given this further assumption, the extrapolation of theory failure would have been totally implausible.

The gist of my argumentation against the projective PI may be formulated as a general lesson. Whatever happened to theories in the distant past of science, whether they were mostly refuted or not, whether they were partly stable or not (with respect to entities, structure, success-conferring parts, or whatever), should have no bearing on the evaluation of our current best theories. Even if, for example, *all* theories accepted as well-confirmed in 1850 had been completely refuted by 1900, that should have no implications for our attitude towards our current best theories (because, once again, the latter enjoy entirely different levels of success today). For the purpose of assessing the truth of our current best

[56] A version of an optimistic induction using formal Bayesian methods is developed in Jan Sprenger (2015).

[57] I am not claiming that categories MOD and HI are empty today, or that they contain fewer theories than category VHI today. More generally, I am neither claiming that science has come to an end, nor that it will end soon, nor that it will ever end. Although I think that, taken together, our current best theories form a fairly comprehensive and very stable world view, it is obvious that our knowledge still has many gaps. Many important questions have not received an answer yet, e.g., questions concerning diseases, natural disasters, hangovers, the future in general, and fundamental levels of physics, and some questions may never receive an answer, e.g., questions about consciousness or global features of the universe.

[58] The optimistic induction can also be used to rebut Stanford's New Induction (2006), at least if one makes the unproblematic assumption that if the conclusion of the New Induction is true (many of our current best theories have "equally well-confirmed" unconceived alternatives), then many or most of our current best theories are false. Needless to say this issue requires more attention than I can give it here.

theories only the last few decades are relevant, while the distant past of science simply does not matter.

Some caveats are in order, though. My defence of the main thesis is clearly just an outline. A lot of work remains to be done such as filling in many details, solving various problems, and addressing several objections. One of the more pressing issues is that the picture of the history of science presented here seems to imply an unwelcome "exceptionalism" about the recent past of science. This exceptionalism requires scrutiny.[59] How, for example, can it be reconciled with the old idea of a steady accumulation of scientific knowledge over the last few centuries? Many realists will presumably think that this old idea has more than a kernel of truth in it and should not be entirely precluded from informing our picture of science. This is only one of a number of important issues that I could not address here (but which I think *can* be addressed). Another one concerns coming to grips with all the idealizations in the very general picture of the history of science presented here, a history which is a convoluted mess of many different scientific fields and disciplines developing at many different speeds and in many different ways.[60] Further issues concern the role of auxiliaries in theory testing; the distinction between data and phenomena (Bogen and Woodward 1988); additional support from the coherence of theories from different fields with overlapping domains (Park 2011); the theory-ladenness of judgements about the amount, diversity, and precision of data, and so on. Finally, my approach faces a sizable number of objections (e.g., Wray 2013), which I cannot discuss in this paper, but have to leave for other occasions. Still I think enough material has been presented to show that the picture of science sketched here captures some very strong overall trends in the history of science and goes a long way to undermine the PI.

At this point, we may finally note a major dialectical advantage that the realist enjoys. The burden of proof to show that the pessimistic induction works lies plainly on the side of the anti-realist (compare Devitt 2011, Roush 2009). She wants to offer an argument against realism, hence she has to get the projective PI going. This means that instead of the realist having to answer the question "Why think that the difference between past and present success is *too big* for the extrapolation of theory failure to be justified?" the anti-realist has to answer the question: "Why think that this difference is *still small enough to permit* the extrapolation of theory failure?"[61] The realist need only provide reasons for doubting that this question can be answered in a positive way, and the empirical material offered in this paper does that in abundance. So, when I formulated and defended a positive claim about the history of science, in the form of the main thesis, I did more than I have to do to defend realism against the projective PI. I did so, because I need the main thesis for a second purpose in addition to arguing against the PI, namely to revive the NMA as an argument in support of realism, to which I will now turn.

5.3 Reply to the objection from ad-hocness

Let us now complete the defence of realism by responding to the objection from ad-hocness. This objection is directed against the move of the realist to restrict realism to theories that enjoy at least as much empirical success as enjoyed by our current best theories. It states that restricting realism in this way is ad hoc, done only for the purpose of saving realism for current theories and not independently justified. Degrees of success will keep rising in the future, or so we can assume. Claiming that it is precisely now, and not at some future time, that we reach a level of success sufficient for truth is not supported by any independent reasons. Invoking the NMA does not help, since doing so is ad hoc and unsupported in the same way.

The reply to this objection is now straightforward. The main thesis shows that the whole situation is quite different from how it was envisaged when the objection from ad-hocness was made. The degrees of success of our current best theories, as measured by the amount, quality, and diversity of evidence in their support, are very high. Therefore, the NMA applied to theories with current levels of

---

[59] This issue was pressed by a referee of this journal.

[60] These concerns were raised by the other referee of this journal. For some useful methodological remarks see Peter Godfrey-Smith (2008, p. 147/8).

[61] The anti-realist either has to show that current levels of success are still below the threshold of degrees of success beyond which past theory failures should not be projected, or she to show that such a threshold does not exist at all, i.e., past refutations imply that theory change will go on forever.

success is extremely plausible. Our current best theories are extremely well confirmed and their success would really be a miracle, if they were false. Furthermore the main thesis implies that all refuted theories were far less successful than the best theories today. Hence, the NMA applied to past levels of success has far less strength, so the success of historically refuted theories need not be considered miraculous.[62] In sum, it is not ad hoc to claim that the NMA strongly supports the inference to the truth of theories with current levels of success, while not supporting (or only rather weakly supporting) the inference to the truth (or partial truth) of theories enjoying the levels of success of the historical theories.

Against this reply the anti-realist may react by posing a really deep question: What, at bottom, is the rational basis for judgments of the sort that, given all the empirical material I presented, current levels of success are “very high” so that the NMA applies with full force to them, and the levels of success of past refuted theories were “moderate”, so that the NMA only applies with little force to them? How are such judgments to be justified in the first place?

These are hard questions. But notice their proper home. They belong to the most central questions about inductive inference and the confirmation of theories. Notice, in particular, that they would have arisen even if there had been no instances of theory change at all. As soon as one recognizes the existence of degrees of success and wants to formulate a realist position, one is faced with the question of which degrees of success suffice for the inference to truth and which don’t. Realists believe, of course, that, at least in central cases, such questions have objective answers that are favourable to realism. One might think that a theory of induction or confirmation provide us with systematic answers to these questions, but no broadly accepted theory has been forthcoming yet and we largely have to rely on intuitive judgments. That is what I did. I relied on the no miracles intuition and intuitive judgments about the confirmational value of good-making features of evidence such as diversity and precision. Therefore, in reply to the above questions I can only suggest that the reader review all the material presented in this paper, to re-examine all the data and its good-making features, and reach judgments (guided by the no-miracles intuition) about the strength of confirmation of past refuted theories and of our current best theories respectively. Hopefully, the judgments of sensible readers will agree with mine.

6 Conclusion

The aim of the paper was to defend scientific realism from the pessimistic induction (PI). I started with a definition of realism according to which it licences the inductive inference from the success of a theory to its approximate truth. This inference is threatened by the PI, which provides examples of theories from the history of science that were successful, but false. To defend realism against the PI I introduced a graded notion of success and modified realism so that the inference to truth is asserted only for theories that are at least as successful as our current best theories, and not for less successful theories, including refuted theories from the history of science that drive the PI. However this version of realism still seems to be threatened by a version of the PI, which I called the “projective PI”, according to which theory failure should be extrapolated along degrees of success from past to current levels of success. I then set out to construct an argument to rebut the projective PI.

At the centre of my argument against the projective PI stood the main thesis of the paper: The degrees of success of our current best theories are very high today, far higher than the moderate degrees of success enjoyed by any of the refuted theories. To support the main thesis I showed, first, that over the last few decades there has been an explosion of scientific evidence, i.e., a huge increase in the amount, quality, and diversity of scientific evidence in many scientific areas. Second, that many theories have benefited from the explosion of scientific evidence gaining huge increases in their degrees of success. And third, that practically none of these theories were refuted. These theories are our current best theories. They are supported by extremely good evidence, practically all of which has been gathered in the last few decades (typically long after the theories were originally accepted). By contrast,

[62] Most refutations were merely partial refutations anyway, if there is anything to all the efforts by realists to show that important parts of successful-but-refuted theories were retained in successor theories.

practically all refuted theories, both of earlier and more recent times, were only supported by moderately good evidence at best. This establishes the main thesis. The main thesis implies that the projective PI is not a sound argument. To sum it all up, whereas past theories were constantly threatened by scientific revolutions, our current best theories have experienced big boosts to their empirical success in the last few decades, have been entirely stable in that time, and are therefore almost certainly safe from scientific revolutions.

References

Adler, J. (2014). "The Science of Society", *Pacific Standard* April 28.

Barrett, J. A. (2003). "Are Our Best Physical Theories Probably and/or Approximately True?", *Philosophy of Science 70*, pp. 1206-1218.

Barrett, Jeffrey Alan (2008). "Approximate truth and descriptive nesting."*Erkenntnis* 68 (2):213 - 224.

Beisbart, C. (2009). "Can we justifiably assume the Cosmological Principle in order to break model underdetermination in cosmology?", *Journal of General Philosophy of Science 40*, pp. 175-205.

Bird, A. (2000). *Thomas Kuhn*, Chesham: Acumen and Princeton, NJ: Princeton University Press.

Bird, A. (2007). "What is Scientific Progress?", *Noûs 41* (1), pp. 64–89.

Bixby, R. (2002). "Solving Real-world Linear Programs: a Decade and more of Progress", *Operations Research 50* No. 1, 1-2, pp. 3–15.

Bogen, J. and Woodward, J. (1988). "Saving the Phenomena", *Philosophical Review 97*, pp. 303-352.

Boyd, R. (1983). "On the Current Status of the Issue of Scientific Realism", *Erkenntnis 19*, pp. 45-90.

Boyd, R. (1990). "Realism, Approximate Truth and Philosophical Method", in Savage, W. (ed.), *Scientific Theories*, *Minnesota Studies in the Philosophy of Science 14*, Minneapolis: University of Minnesota Press.

Brusatte, S. L., Norell, M. A., Carr, T. D., Erickson, G. M., Hutchinson, J. R., Balanoff, A. M., Bever, G. S., Choiniere, J. N., Makovicky, P. J., and Xu, X. (2010). "Tyrannosaur Paleobiology: New Research on Ancient Exemplar Organisms", *Science 329* (5998), p. 1481.

Butterfield, J. (2012). "Underdetermination in Cosmology: An Invitation", *Aristotelian Society Supplementary 86* (1), pp. 1-18.

Bryson, B. and Roberts, W. (2003). *A short history of nearly everything* (Vol. 33). New York: Broadway Books.

Carrier, M. (1993). "What is Right with the Miracle Argument: Establishing a Taxonomy of Natural Kinds", *Studies in the History and Philosophy of Science 24* (3), pp. 391-409.

Chakravartty, A. (2007). *A Metaphysics for Scientific Realism: Knowing the Unobservable*, Cambridge: Cambridge University Press.

de Solla Price, D.J. (1963). *Little Science, Big Science*, New York: Columbia University Press.

Devitt, M. (1997). *Realism and Truth*, 2nd edn, Princeton: Princeton University Press.

Devitt, M. (2008). "Realism/Anti-Realism", in Psillos, S. and Curd, M. (eds.), *Routledge Companion to the Philosophy of Science*, London: Routledge, pp. 224-235.

Devitt, M. (2011). "Are unconceived Alternatives a Problem for Scientific Realism?", *Journal for General Philosophy of Science 42*(*2*), 285-293.

Djorgovski, S.G., A. Mahabal, A. Drake, M. Graham, and C. Donalek, (2012) "Sky Surveys," in *Astronomical Techniques, Software, and Data* (ed. H. Bond), Vol. 2 of Planets, Stars, and Stellar Systems (ser. ed. T. Oswalt), Berlin: Springer Verlag.

Dobzhansky, T. (1973). "Nothing in biology makes sense except in the light of evolution", *American Etiology Teacher 35*, pp. 125-129.

Doppelt, G. (2007). "Reconstructing Scientific Realism to Rebut the Pessimistic Meta-induction", *Philosophy of Science 74*, pp. 96-118.

Earman, J., Roberts, J. (1999). "Ceteris paribus, there is no problem of provisos." *Synthese*, 118(3), 439-478.

Edwards, P. N. (2010). *A Vast Machine. Computer Models, Climate Data, and the Politics of Global Warming*, Cambridge: MIT Press.

Egg, Matthias. "Expanding Our Grasp: Causal Knowledge and the Problem of Unconceived Alternatives." *The British Journal for the Philosophy of Science* 67.1 (2016): 115-141.

Eronen, M. I. (2015). "Robustness and reality" *Synthese, 192*(12), 3961-3977.

Fahrbach, L. (2009). "The Pessimistic Meta-Induction and the Exponential Growth of Science", in: Alexander Hieke and Hannes Leitgeb (eds). *Reduction and Elimination in Philosophy and the Sciences.* Proceedings of the 31th International Wittgenstein Symposium.

Fahrbach, L. (2011a). "How the Growth of Science Ended Theory Change", *Synthese*, 180 (2):139 - 155

Fahrbach, L. (2011b). "Theory Change and Degrees of Success", *Philosophy of Science* 78 (5):1283-1292.

French, S. and Ladyman, J. (2003). "Remodelling Structural Realism: Quantum Physics and the Metaphysics of Structure", *Synthese 136*, pp. 31-56.

Godfrey-Smith, P. (2008). "Recurrent transient underdetermination and the glass half full" *Philosophical Studies*, 137(1), 141-148.

Goodwin, W. (2012). "Experiments and Theory in the Preparative Sciences", *Philosophy of Science 77*, pp. 429-447.

Goodwin, W. (2013). "Sustaining a Controversy: The Non-classical Ion Debate", *The British Journal for the Philosophy of Science 64* (4), pp. 787-816.

Hacking, I. (1983). *Representing and Intervening*, Cambridge: Cambridge University Press.

Holdren, J. P., Lander, E., Varmus, H. and Schmidt, E. (2010). *Report to the President and Congress. Designing a Digital Future: Federally Funded Research and Development in Networking and Information Technology*. online

Hoyningen-Huene, P. (1993). *Reconstructing Scientific Revolutions: Thomas S. Kuhn's Philosophy of Science*, Chicago: University of Chicago Press.

Humphreys, P. (2004). *Extending Ourselves: Computational Science, Empiricism, and Scientific Method*, Oxford: Oxford University Press.

Kitcher, P. (1993). *The Advancement of Science*, New York: Oxford University Press.

Krieger, K. (2006). "All Things being equal", *NEW SCIENTIST 189* (2541), pp. 42-43.

Kuhn, T. (1962). *The structure of scientific revolutions*, Chicago: University of Chicago Press.

Kuipers, T. A. (2000). *From instrumentalism to constructive realism*, Dordrecht: Kluwer Acadamic Publishers.

Kukla, A. (2001). "Theoreticity, underdetermination, and the disregard for bizarre scientific hypotheses." *Philosophy of Science 86* (1), pp. 21-35.

Kukla, A. (2010). "Exceeding Our Grasp: Science, History, and the Problem of Unconceived Alternatives, by P. Kyle Stanford", *Mind 119* (473), pp. 243-246.

Ladyman, J. and Ross, D. (2008). *Every Thing Must Go: Metaphysics Naturalized*, Oxford: Oxford University Press.

Lange, M. (2002). "Baseball, Pessimistic Inductions, and the Turnover Fallacy", *Analysis 62*, pp. 281-5.

Laudan, L. (1981). "A Refutation of Convergent Realism", *Philosophy of Science 48* (3), pp. 19-49.

Leakey, R. and Lewin, R. (1996). *The Sixth Extinction: Patterns of Life and the Future of Humankind*, New York: Doubleday and Company.

Leplin, J. (1997). *A Novel Defence of Scientific Realism*, Oxford: Oxford University Press.

Lewis, P. (2001). "Why The Pessimistic Induction Is A Fallacy", *Synthese 129*, pp. 371-380.

Lupski, J. R. (2010). "Human genome at ten: The sequence explosion", *Nature 464*, pp. 670-671.

Lyons, T. D. (2002). "Scientific Realism and the Pessimistic Meta-modus Tollens." In *Recent Themes in the Philosophy of Science: Scientific Realism and Commonsense*, ed. Steve Clarke and Timothy D. Lyons, 63–90. Dordrecht: Kluwer.

McMullin, E. (1984). "A Case for Scientific Realism" in Leplin, J., ed. *Scientific Realism*, Berkeley: University of California Press, pp. 8-40.

Musgrave, A. (1988). "The Ultimate Argument for Scientific Realism", in Nola, R. (ed.), *Relativism and Realism in Science*, Dordrecht: Kluwer Academic Publishers.

Niiniluoto, I. (1999). *Critical Scientific Realism*, Oxford: Oxford University Press.

Nordhaus, W. D. (2002). *The Progress of Computing*, Version 5.2.2. online

Norton, J. (2008). "Must Evidence Underdetermine Theory?", in Carrier, M., Howard, M., D. and Kourany, J. (eds.), *The Challenge of the Social and the Pressure of Practice: Science and Values Revisited*, Pittsburgh: University of Pittsburgh Press, pp. 17–44.

Park, S. (2011). "A Confutation of the Pessimistic Induction", *Journal for General Philosophy of Science 42* (1), pp.75-84.

Pitts, J. B. (2011). "Permanent underdetermination from approximate empirical equivalence in field theory: Massless and massive scalar gravity, neutrino, electromagnetic, Yang–Mills and gravitational theories." *The British Journal for the Philosophy of Science*, *62*(2), 259-299.

Psillos, S. (1999). *Scientific Realism: How Science Tracks Truth*, New York and London: Routledge.

Psillos, S. (2009). *Knowing the structure of nature: Essays on realism and explanation*. Palgrave Macmillan.

Putnam, H. (1975). *Philosophical Papers I*, Cambridge: Cambridge University Press.

Rainville, S. et al (2005). "World Year of Physics: A direct test of $E = mc^2$" *Nature 438* (7071), pp. 1096-1097.

Roush, S. (2009). "Optimism about the Pessimistic Induction", in Magnus, P. D. and Busch, M. (eds.) *New Waves in Philosophy of Science*, Palgrave MacMillan.

Ruhmkorff, S. (2011). "Some Difficulties for the Problem of Unconceived Alternatives", *Philosophy of Science 78*(5), 875-886.

Ruhmkorff, S. (2013). "Global and Local Pessimistic Meta-inductions", *International Studies in the Philosophy of Science 27*(4), pp. 409-428.

Saatsi, J. and Vickers, P. (2011). "Miraculous Success? Inconsistency and Untruth in Kirchhoff's Diffraction Theory", *British Journal for the Philosophy of Science 62* (1), pp. 29-46.

Speake, C. (2007). "Physics: Gravity passes a little test", *Nature 446 (7131*), pp. 31-32.

Sprenger, J. (2015). "The probabilistic no miracles argument", *European Journal for Philosophy of Science*, 1-17.

Szalay, A. and Gray, J. (2006). "2020 Computing: Science in an exponential world" *Nature 440*, pp. 413-414.

Schummer, J. (1997). "Scientometric Studies on Chemistry I: The Exponential Growth of Chemical Substances 1800-1995", *Scientometrics 39*, pp. 107-123.

Stanford, P. K. (2006). *Exceeding Our Grasp: Science, History, and the Problem of Unconceived Alternatives*, Oxford: Oxford University Press.

Starrfelt, J., & Liow, L. H. (2015). "How many dinosaur species were there? Fossil bias and true richness estimated using a Poisson sampling model (TRiPS)". *bioRxiv*, 025940.

Sullivan, D.B. (2001). "Time and Frequency Measurement at NIST: The first 100 Years", *Frequency Control Symposium and PDA Exhibition, 2001. Proceedings of the 2001 IEEE International*, pp. 4-17.

Tannehill, J., Anderson, D. and Pletcher, R. (1997). *Computational Fluid Mechanics and Heat Transfer,* CRC Press.

Ushijima, I., Takamoto, M., Das, M., Ohkubo, T., & Katori, H. (2015). Cryogenic optical lattice clocks. *Nature Photonics*, 9(3), 185-189.

van Fraassen, B. (1980). *The Scientific Image,* Oxford: Oxford University Press.

Vickers, P. (2013). "A Confrontation of Convergent Realism". *Philosophy of Science,* 80(2), 189–211.

Vickery, B.C. (2000). *Scientific Communication in History,* Lanham, Maryland: The Scarecrow Press.

Wang, S. and Dodson, P. (2006). "Estimating the diversity of dinosaurs", *Proceedings of the National Academy of Sciences 103.37* pp. 13601-13605.

Worrall, J. (1989). "Structural realism: The best of two worlds?", *Dialectica 43*, pp. 99-124.

Wray, K. B. (2011). *Kuhn's evolutionary social epistemology*. Cambridge University Press.

Wray, K. B. (2013). "The pessimistic induction and the exponential growth of science reassessed", *Synthese 190* (18), pp. 4321-4330.

Wray, K. B. (2015). "Pessimistic Inductions: Four Varieties." *International Studies in the Philosophy of Science,* 29(1), 61-73.

Yudkowsky, E. (2007). "Introducing the Singularity: Three Major Schools of Thought", http://www.singinst.org/summit2007/audio/ss07-eliezeryudkowsky1.mp3.