Scientific Revolutions and the Explosion of Scientific Evidence

Ludwig Fahrbach
Preprint

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 both 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[1]

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 threaten to undermine scientific realism.

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[2]. Much effort has then been invested into showing that most instances of theory change of the past were not that radical, in that the parts of the theories that make them partially true survived.

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 (sections 5.1, 5.2, and 5.5). The argument for the main thesis is developed in sections 3 and 4. Section 6 provides a conclusion.

---

[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, Helmut Pulte, Juha Saatsi, Eric Schließer, Gerhard Schurz, Mark Siebel, Matthias Unterhuber, and audiences in Bochum, Düsselforf, Dortmund, Dublin, Oldenburg, Toronto, and Bern. For comments on earlier drafts I would like to thank Claus Beisbart, James Nguyen, Sungbae Park, Sam Ruhmkorff, Paul Thorn, and Ioannis Votsis. I am certain I forgot some people, for which I apologize.

[2] 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)

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 valid.[3] 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. I use the terms "success", "predictive success", and "empirical success" interchangeably. If a theory makes false predictions, it does not count as successful. When realists talk about successful theories, the ones they have in mind are our current best theories. These include 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.[4] 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: "If a successful theory were false, then the success of the theory would be a miracle. 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" (Putnam 1975). This version of the NMA appeals directly to our confirmational intuitions. Other more elaborate versions, which I will not use in this paper, 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, but defending it is not an aim of this paper. Instead I will simply assume that it has probative force.

2.2 The pessimistic induction

Realism is threatened by the pessimistic induction (PI). According to the PI 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 several versions of ether theories. Thus, realists face a situation in which there are two arguments pulling in opposing directions. The NMA supports scientific realism, which the PI threatens to undermine. The two have to be balanced against each other. As they have been presented so far, it is pretty clear that the PI trumps the NMA. What is more, the PI also seems to refute the NMA, because the NMA implies that the predictive success of these refuted (and therefore false) theories is miraculous. So many miracles cannot be tolerated. Hence the NMA fails. Thus the PI seems to refute both realism and the NMA.

The PI undermines realism in the following manner. On most accounts of inductive inference, a statement P *undermines* a statement R, just in case it *supports the negation* of R.[5] We can assume that this is true, if P is the premise of the PI – at least a significant proportion of successful theories from the history of science are false – and R is the realist claim that predictively successful theories are probably approximately true. What is the negation of the realist claim? To simplify things let us identify approximate truth with truth, and count theories that are partly true and partly false, but not approximately true, as false. Furthermore, let us also ignore the difference between the notion of probability

[3] The term "inductive inference" is meant to cover all valid non-deductive inferences.

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

[5] This equivalence is true, 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.

occurring in the formulation of realism and the corresponding relative frequencies. This allows us to take the negation of realism to be the claim that "at least a significant proportion of successful theories are false". Thus understood, the PI is an argument from the premise that at least a significant proportion of successful historical theories are false, to the conclusion that at least a significant portion of successful theories, historical or not, are false. Let us call this argument the "easy PI". So, the conclusion of the easy PI is the negation of realism, and this is how the easy PI undermines realism.[6]

How is the quantifier "a significant proportion" in the premise and conclusion of the easy 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.[7] Furthermore, the NMA is often interpreted to be closely related to IBE (inference to the best explanation). IBE is a 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. In line with this understanding of both the NMA and IBE, realism 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, IBE, and realism face increased 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).

The literature contains, of course, quite a number of different versions of PI.[8] Many versions have conclusions that are stronger than the conclusion of the easy PI. They state, for example, that *most* or *all* successful theories (or current successful theories) are false. However, as we have just seen, in order to refute realism and the NMA, the quantifier of the conclusion of the PI can be fairly weak, "at least a significant proportion" suffices. This does not commit the defender of the PI to the claim that most, or all, empirically successful theories are false, and therefore has the advantage that it is compatible with the fact, sometimes held against other versions of PI, that there have been many theories in the history of science that have been entirely stable for a long time up until the present without ever being refuted (Ernan McMullin 1994, Alexander Bird 2007).

2.3 Two objections

A common response to the easy 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., the former can generally be taken to have made better, and more diverse, correct predictions than the latter. 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 valid 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

[6] Ruhmkorff (2013, p. 409) distinguishes between two kinds of PI: One undermines realism, the other supports the anti-realist statement that some (or most, or most current) successful theories are false. According to the analysis in the text, the premise of the easy PI undermines realism by supporting its negation. I think this analysis also applies to the versions of PI classified by Ruhmkorff, which means that in those cases there is only one kind of PI, not two. Ruhmkorff's distinction between two kinds of PI may then be seen as two ways in which one and the same argument (the easy PI or some other variant) is *used*, either in the manner suggested by my analysis as an argument directed against some version of realism and the NMA in discussions of realism, or else as an argument whose conclusion is an anti-realist position in a dialectic context in which realism and the NMA are not at the centre of attention (even though realism is the negation of the conclusion of the respective form of PI).

[7] Compare Juha Saatsi and Peter Vickers (2011, p. 33).

[8] Ruhmkorff (2013) develops and analyzes an especially convincing local version of PI which concerns certain kinds of medical studies.

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.[9]

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 significant. 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 be refuted." This argument is another version of the PI. I call it "projective PI", because it explicitly extrapolates along degrees of predictive success.[10]

I assume that the projective PI has the conclusion: "At least a significant proportion of our current best theories are false". This conclusion is the negation of realism (in the restricted form), because I assume that, 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".[11] 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 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.[12] Theories have become more successful all the time throughout the history of science and will probably keep 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? As long as the realist is not able to provide an

[9] 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. Some further aspects will be scrutinized towards the end of the paper.

[10] 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.

[11] The projective PI has the form of an inductive inference from a sample of observed cases to the next unobserved case(s) (like the inference from observed ravens to the next unobserved raven). The easy PI has the form of an inductive inference from a sample of observed cases to a whole population (from observed ravens to all ravens). Needless to say, a number of inferentially relevant features, such as the role of degrees of success in the projective PI, are not captured by the two inductive forms.

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

independent reason for restricting the scope of realism to theories enjoying current, but not past, levels of success, his move has to be judged to be ad hoc.

Realist 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 anti-realist can reply that invoking the NMA in this manner 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*. All it tells us is that a level of success *exists* that suffices to infer truth. 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 only pushes the problem one step back: The realist's claim that the NMA already supports realism towards theories at least as successful as our current ones, rather than those at least as successful as future theories is also ad hoc and not based on any independent reasons.

Given the two objections where do we stand? Realism, as restricted to theories at least as successful as our current best theories, 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 easy 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.[13] Fortunately, we need not attempt to decide the matter, because both objections can be fully rebutted, as I will now show.

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*.[14] This is the main thesis of this 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 more distant past, practically all of which enjoyed moderate success at best. Here the term "recent past" is used to denote the last few decades, and the terms "past" and "more distant past" are used interchangeably to denote 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.

[13] 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" to accepted theories from past to present. The New Induction adds a number of novel and interesting considerations to the realism debate, but it arguably ends in stalemate as well (Kukla 2010, Ruhmkorff 2011). What is more, I think it can be countered with the material presented in this paper, as I will briefly indicate below.

[14] Doppelt's (2007) defence of realism against the PI relies on a similar claim, although he seems to use a different notion of success.

Note that the first premise does not talk about refuted theories per se. It just compares our current best theories to the best theories of the 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 more distant past* were merely moderately successful, and the second premise states the same for refuted theories *of the recent past*. The argument for the first premise will take up most of the paper.

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 two terms are chosen in order to highlight the contrast between the two classes of theories mentioned in the main thesis. 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. As we will see, replying to the objection from the past and 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. 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[15] 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 theory’s prediction, 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 significant (or anomalies accumulate), the theory is refuted and a theory change may take place, or so I assume.[16] 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 2). To justify claims that a theory is absolutely (or 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

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

[16] I think these claims about theory testing, as well as similar claims in other parts of the paper, are truisms. They are not shared 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.

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 argument for 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 more 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.

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.[17] Above all, they cannot expect to get recognition for doing so.[18]

[17] 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).

[18] 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. May-

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[19]:

- the Periodic Table of Elements[20]
- 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 more 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

---

be such cases are more wide-spread than it first appears. If so, then all the better.

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

[20] 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.

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 how 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.[21] 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. Therefore, the inference from predictive success to truth is far from trivial, even if the theories are largely or wholly about observables. Hence, 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.[22] 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 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. I want to exclude such theories from my discussion because I think they mount special challenges to realism that require special treatment. Some fundamental physical theories have benefited strongly from the explosion of scientific evidence and are extremely successful today, but there are reasons to doubt that they are approximately true. If they are false, they constitute counterexamples to the inference from current levels of success to approximate truth. There are at least two kinds of cases.

First, fundamental physics seems to offer real-life cases of the underdetermination of a theory by all possible evidence. 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 the very high success of a theory to its approximate truth. For example, put crudely, whether gravitation is a force, a field, or a feature of space-time may not make an observable difference. Also, some of the different interpretations of Quantum Mechanics are observationally indistinguishable.[23] Note that dismissing these examples for the reason that the competing theories don't possess theoretical virtues, such as simplicity or elegance, does not work. They often do possess such virtues to a sufficient extent (that's why I just referred to them as "real-life cases").

Secondly, Quantum Theory and the Special Theory of Relativity have both made numerous true predictions, sometimes with extremely high precision, but the two theories contradict each other. Therefore they cannot both be approximately true, or so one might argue.[24] If this is right, at least one of them is also a counterexample to the inference from very high success to approximate truth. For these two reasons, and possibly more, the inference from the very high success of a theory to its approximate truth has to be qualified in some way when applied to theories in fundamental physics. The upshot is that such theories are a special case in the realism debate, one which needs to be treated separately.

---

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

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

[23] Claus Beisbart (2009) and Jeremy Butterfield (2013) discuss examples of underdetermination in cosmology. If any of their examples involves theories with very high success, then cosmology (which is closely connected to fundamental physics anyway) requires special treatment as well.

[24] Jeffrey Barrett (2003) argues that no clear sense can be made of the claim that these two theories are both approximately true.

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).[25] 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.[26]

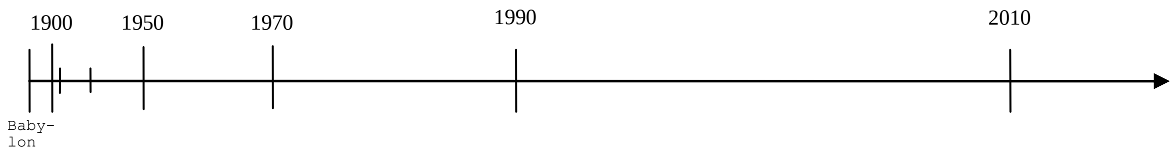

Figure 1. A depiction of the time-line in which the length of any interval is proportional to the amount of scientific research done in that interval. The depiction assumes a doubling rate of 20 years.

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. 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.

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). This sort of growth is typical for the growth of the fossil record in general.[27] One hundred years ago, the number of known fossil genera was in the thousands, today it is 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.

---

[25] De Solla Price (1963) and Vickery (2000). For details and further references see Fahrbach (2009).

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

[27] On tyrannosaur fossils see Brusatte et al (2010, p. 1481), on human fossils see Bill Bryson and William Roberts (2003, Ch. 28), and on early mammalian fossils see the diagram in Fahrbach (2011b).

Secondly, in the very recent past data generation has increased even 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). As a result, the total amount of scientific data is estimated to double every year today (Sazlay and Gray, 2006, p. 413-4).

Consider an example from astronomy. 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 170 million sequences today.[28] In the last couple of years two new databases, TRACE and SRA, overtook GenBank. As of 2010, they harboured ten and 100 times as much data as GenBank.[29] Such a growth is only possible with the heavy use of, and an ever increasing role for, automation.

A final example is the Argo system in oceanography. 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 current 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. Again, the automatic gathering of data has led to huge amounts of data.

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. But not good enough. 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 successful 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.

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.[30] 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

[28] http://www.ncbi.nlm.nih.gov/genbank/statistics

[29] See the charts in James R. Lupski (2010).

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

many kinds of computational problems the number and efficiency of methods and algorithms has much improved in the last few decades.[31]

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.[32] 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.[33]

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. 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. 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. Likewise, 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 was, 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.

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

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

[33] 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).

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.

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.[34]

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. Hence, for many of our current best theories, the relevant data sets exhibit diversity of the deep kind.[35]

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 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 slightly 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

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

[35] 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".

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. 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, or else many of our current best CTs would have accumulated a large number of anomalies, or many significant anomalies, and would have been abandoned at some point, and that's not what we observe. 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 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.[36] As a result, it has passed a large breadth of mainly gentle tests that have provided it with strong confirmation.

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

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.[37]

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 one to two years. By 2013 it had recorded 75 million substances.[38] 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.

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, 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 some theory, then more precise data about the same object or phenomenon are relevant for this theory as well.[39]

Data become more precise when scientists improve upon existing, or develop new, kinds of instruments and measurement techniques.[40] Enormous progress in both has occurred in the last few decades. 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).[41]

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

---

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

[38] http://www.cas.org/news/media-releases/75-millionth-substance

[39] 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.

[40] 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 objection analogous to the projective PI that theory change should be extrapolated along the improvement of method.

[41] 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.)

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[42].

The example shows once again that confirmation en passant can be very strong confirmation. Geologists mostly accepted plate tectonics already in the 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 can be very severe and can often produce big boosts to the success of theories.

The fourth example is more specific, namely a test of Einstein's equation $E = mc^2$ in the year 2005. In this test 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%.[43] 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.[44] 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 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. But 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^{-17}, or one false second per 3 billion years (Chou et al 2010). 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.

---

[42] 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.

[43] 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.

[44] 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.

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 more 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 (e.g., no computers) 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 more 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, theories of the more distant past were only moderately successful at best. Hence, this also holds for all *refuted* theories of the more 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 (Fahrbach 2009, p. 103).

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 more 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 more distant past supports the claim that there are no RATs. Furthermore I have kept an eye out for theories 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). 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.[45] 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 requires careful elaboration which I cannot provide here, hence I will only briefly sketch it and not use it outside this sub-section.

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.[46]

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 pat-

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

[46] 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.

tern: 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.[47] 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.[48] 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.[49] 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 more 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 theories only the last few decades are relevant, while the more distant past of science simply does not matter.

Some caveats are in order. What I presented in this paper is clearly just an outline that needs to be developed further. For example, I determined the degrees of success of the refuted theories of the more distant past by relying entirely on general considerations without actually examining any of these theories. This is only one of a number of important issues that I could not address. Others 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 disciplines or sub-disciplines (Park 2011); the theory-ladeness of judgements about the amount, diversity, and precision of data, and so on. There are also various objections that need to be discussed. I will discuss two in the next section, while the rest have to be left for other occasions.

5.3 The objection from relativity

In this and the next sub-section I discuss two counterarguments against my rebuttal of the projective PI. I think both are hopeless, but discussing them will prove instructive.

The first counterargument, which I call the objection from relativity, goes as follows. The main thesis states that our current best theories are far more successful than the refuted theories of the past.

[47] See, e.g., Stanford (2006, Ch 1), Norton (2003), Kukla (2001).

[48] The relationship between the optimistic induction and the conclusion of the projective PI is analogous to the relationships between the easy/projective PI and unrestricted/restricted realism, as analyzed in sections 2.2/ 2.3.

[49] 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.

This judgment is obviously made from the perspective of current science. However, relative to future science, say the science of 2100, this judgement seems to be inaccurate, because the amount, and quality, of evidence will probably keep growing exponentially. Instead, from the perspective of 2100 it seems plausible to judge that the best theories of today are only a little more successful than the refuted theories of the past, while the really big increase in success will come in the future between today and 2100. So, what for us today looks like a big difference in success will look like a small or tiny difference from the perspective of scientists in 2100 (Figure 2). But if the difference in success between our past and our present is small or tiny, then the extrapolation of theory failure goes through after all, and my rebuttal of the projective PI does not work.

What this seems to show is that it is possible to use different scales to assess differences between degrees of success in different ways. One scale is relative to current science and its highest levels of success, another scale is relative to the science of the year 2100 and its highest levels of success, and relative to other times still other scales may be chosen.[50] The question then is which of these scales is the correct one, and how its correctness can be shown. The objection from relativity states that there is no objective way to choose between the different scales. There is no reason to suppose that one scale is superior to all others. Every community of scientists can make judgments about which differences in degrees of success allow an extrapolation of theory failure and which don't from the point of view of their time. But beyond that no such judgments can be justified on objective grounds, at the expense of the corresponding judgments relative to other times. In other words, no scale is better than any other.

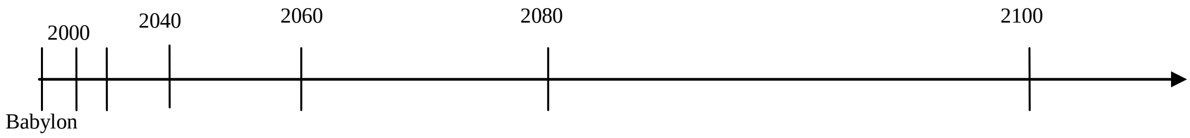

Figure 2. The growth of scientific research as seen by scientists of the year 2100, assuming a doubling rate of 20 years.

My reply to the objection from relativity is that such a relativity of scales is entirely unacceptable. It is certainly unacceptable for realists. Generally speaking which differences in degrees of success allow an extrapolation of theory failure and which don't is a certain kind of confirmational question. As such it does not differ fundamentally from other kinds of confirmational questions, such as which sets of evidence make a hypothesis believable and which don't. If one thinks that the latter questions have objective answers, then it is plausible to think that the former questions do too. Realists will typically think (in line with their acceptance of the NMA) that the latter questions do have objective answers, at least for central cases. Hence they will hold that questions about the extrapolation of theory failure have objective answers, and that, in particular, the question whether the projective PI (the extrapolation of theory failure from our past to our present) is valid or not has an objective answer. So, realists will reject the relativity of scales.

But the relativity of scales is unacceptable to the anti-realist as well. If it were correct, the projective PI would be valid relative to some times, e.g., relative to the year 2100, while invalid relative to other times, e.g., relative to today, and beyond that no objective judgments could be made. But an argument cannot be relative in this way. Either an argument is valid or it is not. Hence, when the anti-realist presents the projective PI, he himself assumes that not all scales are equal and that objective judgments about the extrapolation of theory failure are possible.

[50] Scientific research has generally been growing exponentially over the whole history of science since the Scientific Revolution. Therefore, scientists at all times since then could have made, and often did make, the observation that most of the increase in scientific research (and scientific manpower and number of publications) took place in their respective recent past. Hence, it seems plausible (but also deserves further historical investigation) that the same holds, by and large, for the increase in amount and quality of evidence and the increase in degrees of success of the best theories.

To drive home the point let me describe a scenario about which most people can agree that the extrapolation of theory failure is highly implausible. Imagine a really huge difference in empirical success. Imagine, for example, that in the year 2200, scientists will have hundreds of completely different kinds of telescopes and microscopes; will be able to measure time, distances, masses, and many other quantities with a precision of dozens of decimal places; will have sequenced the genome of every kind of organism on earth; and so on. The result would be an almost unimaginable amount of evidence of extremely high diversity and quality. In such a situation can one extrapolate theory failure from the degrees of success of the refuted theories of our past to the degrees of success of the best theories of the year 2200? I take it that most realists, and probably many antirealists, would agree that such an extrapolation would be utterly implausible. Hence differences of success exist for which an extrapolation of theory failure is entirely unbelievable. Furthermore, let us imagine that after 2200 science keeps growing exponentially. Then the difference in success between our past and 2200 will look huge for us, big for people in 2200, and tiny from the perspective of people of 2400, for instance. Still, the fact that such a variety of different outlooks are possible should have no bearing on the objectivity of the judgment that in this scenario the difference in success between our past and the year 2200 will be (timelessly) too big for the extrapolation of theory failure. Thus, we can conclude that the objection from relativity fails.

The consideration of the last paragraph shows that, in principle, differences in success exist over which theory failure should not be extrapolated. But our concern is with the difference in success between *our* present and *our* past. The anti-realist asks: (Q) Why think that *this* difference is too big for the extrapolation of theory failure to be justified? Now I have, of course, provided ample empirical material about the explosion of scientific evidence and its relation to the degrees of success of scientific theories culminating in the main thesis that was meant to answer this question. But maybe the anti-realist is not satisfied with this answer, because he thinks my efforts so far only give rise to deeper questions, questions that may have bothered some readers all along. We want to make objective judgments of the sort that some differences in degrees of success allow the extrapolation of theory failure, while others don't. But what, in the end, could be the rational basis for such judgments? How exactly does all the empirical material I presented support the claims of the main thesis that current levels of success are "very high", while the levels of success of refuted theories were "moderate at best", so that the difference in success is too big for the extrapolation of theory failure?

My reply is as follows. One might hope that a theory of induction or confirmation provide us with answers to these questions. Unfortunately we do not yet have such agreed upon theory. We have some approaches, e.g., formal approaches such as Bayesianism, but they are far from widely accepted, come in many different varieties, and have actually little to say about these questions so far.[51] Therefore, we largely have to rely on our intuitions. That is what I did when arguing for the main thesis.

Actually, at this point we may finally note a significant dialectical advantage that the realist enjoys. The burden of proof to show that any form of pessimistic induction works lies plainly on the side of the anti-realist. She wants to offer an argument against realism, hence she has to get the PI going (Devitt 2011, Roush 2009). Therefore, instead of the realist having to answer question (Q), the anti-realist has to answer the question: Why think that the difference between past and present success is *still small enough to permit* the extrapolation of theory failure? 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 abundantly. A similar dialectic shift affects the deeper questions following question (Q) (about a systematic justification or grounding of judgments of the sort that such-and-such differences in degrees of success allow the extrapolation of theory failure, while others don't). It is again the anti-realists who, when presenting the PI, should provide such a grounding. They have not offered any so far, relying instead on intuitive judgments. Hence I can do likewise when arguing against the PI.

5.4 People in the past again

The second counterargument is similar in structure to the objection from the past which gave rise to the projective PI. It runs as follows. People at some point in the more distant past, i.e. more than some decades ago, could have reasoned in the same way as I did, namely that the difference in success

[51] Roush (2009) develops some ideas for a theoretical grounding of the PI.

between *their* past and *their* present was likewise too big for the extrapolation of theory failure. But then their theories were refuted nonetheless. Hence, their reasoning would have failed. The difference in success between their past and their present was small enough to extrapolate theory failure. Since our situation today is not essentially different from the situation of these people, my reasoning against the projective PI will also fail. Hence, the difference in success between our past and our present is likewise not too big to extrapolate theory failure.

This argument has several problems. Let me just point out two. The main thesis assets that there has been a period of stability during which no best theory of any scientific discipline was refuted at any time during that period. So, people at some point in the past could have only reasoned in the same way, if they likewise had observed in their past that theory change had stopped for some time simultaneously for all best theories ("best" according to their time). It is not clear that such a period of simultaneous stability exists in the more distant past of science. Laudan's examples of refuted theories, for instance, seem to be distributed fairly evenly over times prior to 1900.

But let us assume for the sake of argument that anti-realists succeed in finding a period of simultaneous stability.[52] To be concrete, let us assume that such a period ended in 1900. Then a deeper problem with the argument is the following. The argument assumes that the situation in 1900 is essentially the same as our situation today. This assumption is used twice. First, it is used to transfer my reasoning from the present to the year 1900, i.e., to claim that people in 1900 could have reasoned in the same way as I did (that the respective difference in success is too big for the extrapolation of theory failure). This is fine. But then the assumption is used a second time to transfer the fact that the reasoning of people in 1900 would have failed back from 1900 to the present. This transfer is problematic. The basis of the failure of the reasoning in 1900 is the theory refutations that occurred after 1900. They are responsible for the fact that the reasoning in 1900 failed, they are the "negative force". Hence, the projection of the failure of the reasoning from 1900 to the present is tantamount to the projection of theory failure itself from 1900 to the present. But the legitimacy of the latter projection, that the difference in success between the past and the present is small enough to justify the extrapolation of theory failure, is precisely what this counterargument aims to show. In other words, the argument has to assume as a premise a statement that amounts to the very statement that is its conclusion. Hence, it is begging the question.

We can shed more light on this problem as follows. Think of the sequences of best theories of all successful scientific disciplines over the whole history of science up until the present. In the more distant scientific past it exhibits a certain pattern of (partial) theory change, which was followed by a period of stability in the recent past. Then *whatever* the pattern of theory change in the more distant past of science, it is always an additional question how far the negative features of this pattern are to be projected along time (or along degrees of success) to some given later point in time. All the empirical material I presented was meant to address this question, i.e., was meant to show that the difference in success between past and present is actually too big to extrapolate any negative features of that pattern from past to present. In contrast, the counterargument does not address this question at all. Instead it simply assumes that "our situation today is not essentially different from the situation of the people in the past."

To drive home the point think again of a future time for which it is entirely clear that the difference in success between our past and this future time is too big to justify the extrapolation of theory failure. Assume, for example, that this is true for the above scenario, in which evidence keeps growing exponentially until the year 2200. Apply the counterargument to such a scenario. Then the flaw becomes evident: The counterargument aims to show that the difference is small enough to extrapolate theory failure from our past to the year 2200, in spite of what we just assumed about the scenario. But it has to assume as a premise what amounts to the very same claim, namely that the situation in 2200 is not essentially different from our past.

5.5 Reply to the objection from ad-hocness

So, the projective PI fails; realism is not threatened by past refuted theories. But we still have to deal with the objection from ad-hocness. This objection is directed against the move of the realist to restrict

[52] Such a period of simultaneous stability of the best theories across many scientific fields or disciplines, followed by a stretch of renewed theory change, would certainly be a curious phenomenon.

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 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.[53] 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 produce counterarguments similar in kind to those presented in the previous two sub-sections. However, the ensuing discussion would not produce any new arguments. Hence, I limit myself to just one set of questions which parallel those asked earlier: 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. Once again, one might think that a theory of induction or confirmation provide us with systematic answers to these questions, but, once again, 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, I can only suggest that the reader review all the material presented above, to re-examine all the evidence 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 who 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". Ac-

[53] 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.

cording to the projective PI, 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, 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 and "paradigm shifts", 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 and "paradigm shifts".

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.

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.

Chou, C. W., Hume, D. B., Koelemeij, J. C. J., Wineland, D. J. and Rosenband, T. (2010). "Frequency Comparison of Two High-Accuracy Al+ Optical Clocks". arXiv: 0911.4527v2 [quant-phy], 2 Feb 2010.

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.

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.

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

Fahrbach, L. (2009). “The Pessimistic Meta-Induction and the Exponential Growth of Science”, in Hieke, A. and H. Leitgeb (eds), *Reduction and Elimination in Philosophy and the Sciences. Proceedings of the 31th International Wittgenstein Symposium, 2008,* vol. 11. Walter de Gruyter, pp. 95-111.

Fahrbach, L. (2011a). “How the Growth of Science Ended Theory Change”, *Synthese 180.2*, pp 139-155.

Fahrbach, L. (2011b). “Theory Change and Degrees of Success” *Philosophy of Science 78*, pp. 1283-1292 .

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

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

Howson, C. (2000). *Hume's Problem: Induction and the Justification of Belief*, Oxford: Oxford University Press.

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.

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.

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

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.
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.
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.
van Fraassen, B. (1980). *The Scientific Image*, Oxford: Oxford University Press.
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. (2013). “Success and truth in the realism/anti-realism debate”, *Synthese 190* (18), pp. 4321-4330.
Yudkowsky, E. (2007). “Introducing the Singularity: Three Major Schools of Thought”, http://www.singinst.org/summit2007/audio/ss07-eliezeryudkowsky1.mp3.