Why Some Sciences Achieve More Rapid Progress

Andrin Spescha
ETH Zurich
KOF Swiss Economic Institute
spescha@kof.ethz.ch

June 1, 2026

Abstract

This paper explores how empirical testing differs across the sciences and how this influences their respective progress. It compares the sciences along two key dimensions. First, sciences working mainly with physical experimental setups can obtain more accurate empirical results than sciences working only with data, since the tangible nature of the experiment allows to better detect and correct errors. Second, sciences studying small physical objects can benefit from empirical testing that is faster, less expensive, and better parallelizable than sciences studying large, not clearly defined systems. The combined ability to quickly generate numerous and accurate empirical results allows for a mutually beneficial interaction between both theory and experiment that makes rapid progress possible. Working with physical technologies more specifically cannot only benefit from this combined ability, but offers important additional advantages in empirical testing, such as a physical manifestation of auxiliary hypotheses, clear signals whether the technology works appropriately, or the capacity to better evaluate corresponding theories. These factors jointly contribute to again more accurate empirical results that also improve coordination among scientists. Consequently, sciences with technologies at their core will experience the fastest progress. The advancements of these technologies in turn provide the foundation for a well-functioning ecosystem linking together science and industry.

Keywords: scientific progress – experiment – trial and error – empirical results – technology - innovation ecosystems

1. Introduction

Science is one of the few human activities where we systematically search for errors and, when we detect them, over time also correct them accordingly. Science learns from the errors it makes. This is why we can say that science makes progress. In most other human activities, we observe change but rarely progress, since gains also bring about losses, and we often have a hard time to even define the change itself (Popper 1963). In fact, a certain field is usually seen as a genuine science if it is successful, that is, if it makes significant progress over time. However, not all sciences show the same rate of progress. For example, Nelson (2008, p.495) argues that: „the knowledge that has been won by the behavioral and social sciences is far less powerful than the knowledge that has been won through the natural sciences". Empirical evidence indeed supports such a "hierarchy of the sciences", where the natural sciences outperform the social sciences on various indicators (see, e.g., Fanelli and Glänzel 2013, Simonton 2004, Lamers et al. 2021). The difference between the natural and the social sciences is probably best illustrated by the practical results that have emerged from them. Consider, for instance, the introduction of modern technologies into our life and contrast this with the introduction of important economic or social policies. Especially the engineering sciences, but also the various life sciences, have delivered much more tangible results here than the social sciences. This paper therefore searches for important factors that cause such differences in progress across the sciences.

The paper defines scientific progress as an increase in the truth content, or verisimilitude, of scientific theories (Popper 1972, Niiniluoto 1987, 2024). A science achieves progress when its theories increasingly correspond to the empirical facts. We can, of course, never be certain whether some theory is indeed true. We may, however, be able to observe whether successions of theories better approximate the empirical world over time. Scientific progress could alternatively also be defined as increase in problems solved (Kuhn 1962, Shan 2019), knowledge (Bird 2007, 2022), understanding (Dellsén 2016), practical dimensions (Mizrahi 2013), or as ultimately subjective (Rowbottom 2023) (for an overview of these different accounts see Shan 2023). However, the precise definition of scientific progress is less central for this paper, as these accounts would likely agree on most of the contrasts made in the paper. For example, they would agree that the natural sciences have in general achieved more progress than the social sciences over the last decades. Some fields will achieve more scientific progress on one account and less on another, but they would likely overlap in their assessment of large differences in progress, such as the strong advance of the computer sciences. The paper at hand emphasizes instead how crucial numerous and accurate empirical evidence is for scientific progress, which depends less on the exact account we choose to frame the arguments. Crucial is only that, over time, sciences benefiting from a faster growing number of empirical facts can rely on stronger constraints that facilitate introducing overall better theories, that is, theories that have higher verisimilitude, solve more problems, lead to more knowledge, enable better understanding, complement practical dimensions, or are valued more by the relevant groups.

Importantly, practicing scientists are not to blame that some scientific fields progress more slowly than others. It is instead the nature of the empirical testing in a scientific field that determines its progress over time. The sciences differ greatly from each other in this respect.

The paper argues first that sciences working mostly with physical experimental setups can better test auxiliary hypotheses and build around them if they are false. In contrast, sciences working mostly with data face difficulties in narrowing down false auxiliary hypotheses in their data collection, processing, and analysis. Consequently, sciences that work with physical experimental setups can produce more true, certain, and detailed empirical results than sciences that work mostly with data. However, the quality of empirical evidence alone is not sufficient; the provided quantity is very important as well. Empirical testing differs crucially between sciences that deal with smaller, simpler, and clearly defined physical objects and sciences that deal with larger, more complex, and less clearly defined systems. The former is more representative of work in laboratories from physics to chemistry to biology, but, importantly, also of the many engineering and applied science disciplines. The latter is more representative of sciences that investigate aspects of economies, societies, or politics. In fields that work with small physical objects, empirical results arrive faster, are less expensive, and allow for parallelization. Testing quickly produces results, the different trials do not cost too much, and they can be replicated in various ways and thus also meaningfully extended. In contrast, in fields that work with large systems, empirical results arrive more slowly, are more costly, and remain more isolated. Taken together, the higher accuracy and better availability of empirical results in sciences that work with small, simple, and clearly defined objects in mostly physical experimental setups allow them to pursue a mutually beneficial interaction between both theory and experiment that leads to in comparison more rapid scientific progress.

Physical objects in the form of technologies, which constitute the core of most engineering and applied sciences and include here both hardware and software, offer a whole series of further aspects fostering scientific progress. First, these sciences often set the experiment within the creation of the technology itself, such that a violation of crucial assumptions becomes more difficult. Second, technologies provide scientists with a clear signal when they work appropriately. Third, the interaction between theory and experiment is especially close in the case of technologies and allows for continuous mutual evaluations. Fourth, the physical existence of the technology forces the various contributions of different scientists to better fit each other. Together these four aspects allow for even more accurate empirical results that in addition also improve coordination among scientists to advance their respective fields. Sciences that interact with, build on, and create technologies can therefore achieve the fastest progress.

The contribution of the paper is thus a novel explanation of why practical sciences like the computer sciences, robotics, or artificial intelligence, which have physical technologies at their very core, can achieve especially rapid progress. They can rely on a distinct set of characteristics of their empirical testing that is favorable to progress and not accessible to other sciences.

The paper is organized as follows: Chapter 2 provides an overview of the most popular theories of scientific progress. Chapter 3 emphasizes trial and error, if applicable, as the most promising tool to achieve progress. Chapter 4 shows this by an investigation of some practical tasks and illustrates how this translates to different sciences. Chapter 5 lays out the six most central characteristics of empirical testing that jointly determine scientific progress. Chapter 6 presents four important advantages for scientific progress that only the technical sciences can profit from. Chapter 7 describes how this in turn influences the organization of science-industry relations. Chapter 8 concludes and makes some predictions about future scientific progress.

2. Theories of scientific progress

The three most prominent theories of how science progresses are probably Popper (1959, 1963), Kuhn (1962), and Lakatos (1978). Together the three provide the most realistic image of science. In many scientific fields, scientists are working on their respective projects by conjectures and refutations, while at the same time they share basic principles in the form of major theories, or a paradigm, with other scientists. Meanwhile, they also participate in the ongoing competition between different theories at the research frontier. Individual scientists thus establish some specific empirical fact or also some more minor theory. Yet the larger theoretical units still develop within different research programs or even successive paradigms. What makes science so successful is a mixture between Lakatos' competition, Kuhn's coordination, and Popper's dialog. At the core, however, stand the individual scientists, who pursue by usually more low-level conjectures and refutations. Behind this process of conjectures and refutations stands in turn the broader principle of trial and error. The next chapters will outline the power inherent to trial and error, which is most clearly visible in many practical tasks such as the construction of new technologies. If a science can build successfully on such technologies, like some of the engineering disciplines, it will develop faster than more basic sciences.

3. Trial and error

Trial and error is often thought to apply only to situations where individuals have little knowledge and just try out all available options until they finally hit upon a specific trial that works out. However, the principle of trial and error has much wider applications. Trials do not necessarily have to be uninformed. A trial can take the form of an elaborate theoretical hypothesis or even an entire scientific theory, as in Popper's conjectures and refutations.

Prior knowledge can go a far way in shaping successful trials. Humans can learn from past trials things that help them generating superior trials. Through experience we can build strategies that allow solving future problems much faster and enable sorting out which trials are effective and which ones lead into dead ends. Such strategies, or heuristics, help reducing the search space of trials (Simon 1996). They use information in the setup of the problem to identify promising search paths. The speed with which a problem is solved thereby depends not on the totality of the search space, but on the quality of the strategies. Overall, this translates into a self-reinforcing process. New and better trials lead to learning that in turn leads to new and better trials. For example, creating any technology relies heavily on all the learning from past trial and error. It provides cumulative knowledge that builds up over time.

However, trial and error does always include at least some "blind" component. When people solve a problem without any trial and error, we say that they already knew the answer. The trial is instead what goes beyond what we already know and into the unknown. Only those trials that go beyond current knowledge can lead to significant changes in our understanding. Blindness begins where existing knowledge ends. This does not mean the trials are random though. It only denotes that the outcome of the trials cannot be foreseen. If it could, the knowledge obtained would not be new (Campbell 1974).

Trials can therefore be intuitive or thought-through, simple or complex, naive or sophisticated, plain or elegant, practice- or theory-inspired. For instance, scientists usually rely on existing scientific knowledge, that is, on previous theories, evidence, or experiences. They use this understanding together with logical inferences to arrive at new trials. This means that there is always an ex-ante selection of trials, and only the more promising approaches are tested. However, how the trials were formed, what they are exactly, and where they originate from is of lesser importance. The more relevant aspect is that the trials can be tested, repeatedly and in various ways. Because if we have the ability to rigorously test trials, we will eventually hit upon a trial that works. Only testing can deliver the answer to whether we were right or not.

4. The power of trial and error

4.1 Practical tasks

Many practical tasks rely heavily on trial and error. For example, in programming computer code, which due to the continuously increasing digitalization has become very important in many areas over the past few decades, trial and error is essential. Software is in many industrial applications now even more important than hardware. For instance, the "brain" of a modern robot is usually a complex optimization problem using data from input sensors, images, and actuators. To control the robot has become the central issue.

Programmers usually build up their code step-by-step, whether self-written or AI-assisted. Every coding trial either runs through or results in error, that is, error messages, test failures, unexpected outputs, etc. These errors guide the programmers throughout. Sometimes the errors are specific, and they can be easily located in the code. Sometimes they are unspecific, and it can take various attempts to locate the source of the problem. The more complex the code becomes, the more this is an issue. The key is that programmers eventually know where they stand. They must correct the errors that interfere with the code's functioning and cannot build reliably on unresolved errors. The process of coding itself forces the programmers to continually exclude falsity. What they have in the end is a tool that they can use.[1] The repeated fixing of errors, or bugs, is not only central in programming but in the construction of any technology. The key is successfully solving all the little problems that can foul up its functioning. Errors must be eliminated or contained one by one until the technology works.

Vincenti (1990) describes how important trial and error has been in the history of aeronautics. The earliest attempts to construct flying machines in France at the beginning of the 20$^{\text{th}}$ century relied heavily on trial and error. The French built various types of machines and proof tested them in flight, while doing little systematic research to guide their attempts. Today, engineers typically design a candidate device and test it by mathematical analysis or experiment until it does the required job. However, this always requires many iterations, with the results of one trial informing the next one, before the final plans can be released for production. Trial and error thus remains important for the construction of advanced airplanes also today. The whole process has become much more systematic though, with for instance the use of wind tunnels, in which „aerodynamic wing design proceeded by a highly structured, time-honored combination of theory and experiment" (Vincenti 1990). Vincenti (1990) argues that such a use of trial and error is universal and characteristic for all branches of engineering.

[1] Today programmers also work in close collaboration with AI agents. Programmers have become the architects that define the goals, requirements, and system design, while the AI agents generate most of the code. The AI agents themselves rely crucially on trial and error, too. They iteratively refine the code based on feedback from different errors until the predefined goals are reached. However, the programmers must still review the written code and check whether it indeed works. Overall, AI agents have greatly accelerated the coding process, as they have shifted it away from writing toward verification.

The successful construction of complex technologies demonstrates that trial and error is a powerful method to advance knowledge. If trial and error is possible, a field will be able to experience steady progress. If it is not possible, a field will progress more slowly. Unfortunately, only some sciences can harness the full power of trial and error. In the next chapter, we will have a closer look at these fields.

4.2 Science

Trial and error is working well not only in practical tasks but also in certain research experiments in science, such as in laboratory biology, where scientists tinker on diverse organisms. „Science is at heart a matter of sophisticated trial and error" (Shiffrin et al. 2018, p.2636), or "a tedious, painstaking process of selective trial and error" (Simon 1977, p.302). The difference in the trial and error process between science and practical tasks is that in the former it is more systematic, slower, and focused on fewer trials.

How does trial and error work in biology, for example? Scientists test a given hypothesis, often intentionally simplified, about how some organism behaves. They then conduct an experiment to observe how the behavior of the organism deviates from the hypothesis. These deviations the scientists use to construct a new, better hypothesis. Scientists subsequently test this hypothesis, too, and so forth. The control over the experimental setup enables scientists to vary specific conditions for exploration.

In trial and error in the laboratory, theory and experiment develop together. They both improve upon each other by providing directions for successively better theories and better experiments. Theory shows where experiment must look, and experiment shows where theory must be corrected. During this process, false alternatives for theory and experiment are eliminated, until finally a specific theory emerges that is well corroborated by specific experiments. However, this process is never finished. Scientists always try to find even better theories and experiments.

Examples for this kind of trial and error are numerous across the sciences. For example, Michael Faraday's discovery of electromagnetic induction (Simon und Klahr 1999), Hans Krebs's discovery of the ureal cycle (Kulkarni and Simon 1988), Jean Perrin's studies of Brownian motion (Mayo 1996), Alexander Fleming's discovery of penicillin (Bud 2007), James Watson and Francis Crick's discovery of the DNA (Watson 1968), or also the laboratory work in experimental biology more generally (Weber 2004). In all these cases, the scientists pursued repeated trial and error to generate and then test sequences of hypotheses until they obtained consistent result.

Importantly, in experimental practice, only some trials will refer to theoretical conjectures. Trial and error takes place in two crucial but separate instances. First, scientists vary the experimental setup, apparatus, or technique to eliminate the influence of background factors and, second, they vary the experimental conditions or treatments to advance their theoretical knowledge. In both cases, learning proceeds step-by-step. Scientists make the treatment exogenous first and then tinker around with it. Hence, many trials will not be theoretically important, but rather concern low-level experimental conjectures, such as excluding possible background factors like the influence of air, light, temperature, etc. The growth of knowledge happens in experimental tasks (Mayo 1996) as well as in theories (Popper 1963).

The many pages long laboratory reports in biology, for example, are evidence for the numerous low-level conjectures in experiments. For established experiments, scientists can go through all the steps in the laboratory report and learn how the experiment should be conducted. In contrast, for new experiments, scientists must find ways how it should be conducted. They need to find a series of successful steps, or trials. Because scientists usually have few theories available at such low levels of the experiment, they often rely on blind trial and error. Previous experience thereby greatly supports finding some path. The result of this process is then again a many pages long laboratory report with the trials that have survived (Knorr Cetina 1999).

4.3 Practical Science

The "linear model" says that basic science is the origin of our knowledge and applied science uses this knowledge to create practical applications. The latter is what creates utility, in for instance new technologies. Applied science could not function without basic science. However, this linear model has been rejected as inaccurate (Douglas 2014). Academic scientists do applied science, and industrial scientists do basic science. Moreover, applied science also leads to the creation of new theories and basic science to new applications. The connections between the two are complex with long lasting feedback processes. Often older insights from science are even more useful for application than the most recent insights.

In practice, pragmatic trial and error may deliver some technological advance, which science then tries to understand through discovering some general principles. This in turn gives inputs for even further technological advance. In those fields where technique and understanding co-evolve, technological advance is fast, as they reinforce each other (Nelson and Nelson 2002). They are both best pursued when there is only little distinction between them. Many technologies can also not be fully tested in the lab. They show their merits only in the markets, where competition will show what is best. History gives us many examples of fields where sequences of practical inventions, developed through trial and error, lead to theoretical analysis, such as the steam engine that led to thermodynamics (Hacking 1983).

The most practical of the sciences, the various engineering disciplines, have their own scientific basis today. They are not just applied sciences. Engineers cannot ignore friction, viscosity, or turbulence when designing an airplane. They had to develop their own methods to construct airplanes, which over time led to the creation of their own theories (Vincenti 1990). Similarly, the construction of plants to produce chemicals created the scientific discipline of chemical engineering. The optimal plant sizes were too large to rely on the in the principles of chemistry and required an own discipline with its own methods and theories (Rosenberg and Steinmueller 2013). In the engineering sciences, knowledge gathered in the interaction with physical objects gets codified and converted into mathematical language. The object of the study is thereby designed. The engineering disciplines create the very objects they themselves analyze. The knowledge of the engineering disciplines is in turn not just used to understand their objects but also used to improve them.

4.4 Simulation

In engineering, the growth of knowledge has made it possible to simulate most trials (Vincenti 1990). Such vicarious trials can be physical (e.g., wind tunnels, pilot plants) or artificial (e.g., computer simulations, analytic calculations). For example, the field of reinforcement learning, which relies heavily on trial and error, has been useful for training the "brain" of modern robots. The robots can learn through computer simulations how to behave in actual environments. Vicarious trials have greatly extended the ability to select promising trials. They are what differ modern engineering from past searches for technology. Vicarious trials allow to just run through many trials without the need for too much creative thinking. Engineers thus need less ingenuity in the formulation of the trials than in science, as they can substitute it by rapid testing. Even though direct trials of the physical prototype will have the last word in each case, vicarious trials can help greatly in suggesting which direct trials to pursue.

4.5 Rapid testing

Hence, we observe trial and error in practical tasks, in science, and in practical science. What crucially differentiates practical work from science is the constant testing. Whereas in science empirical evidence arrives only at long time intervals, in practical tasks such as coding it often arrives within seconds. The testing of physical prototypes is similarly orders of magnitudes faster than most testing in science. Yet only constant testing allows for continuous improvement and thus also for progress, be it in technology or (practical) science.

Of course, progress over numerous trials is seldom constant. Sometimes trials do not lead to any progress or even to regress. Some search paths can turn out to be a dead end. Other times trials lead to very rapid progress. The timeline can be very different for different situations. Progress may happen over hours, days, months, years, or decades. Bursts of progress then sometimes turn into scientific revolutions.

5. Different empirical testing

How can we learn during trial and error? The trial needs to show by how much we are off. We need information on the prediction error, that is, the discrepancy between the expectation and the actual outcome of the trial. This information then allows us to improve the setup for the next trials. In the sciences, this means information on how much an empirical result deviates from the hypothesis or the predicted outcome. A precise falsification can give insights about how to change the theory. In contrast, if the empirical result aligns with the hypothesis or the predicted outcome, the theory can be developed in more detail. In practical tasks, learning from trials means information on how well an object achieves the intended outcome. Deviations from the target provide practitioners with information on how to improve the object. In both cases, science and practical tasks, the outcomes need to be attributable to the trials. Experimenters need to be able to identify, control, and replicate their trials. If, in contrast, they run many trials at once, or the setup is complex, this will become more difficult.

Learning will be especially high if we have a precise measurement of the prediction error. In the sciences, this requires a precise hypothesis and, even more important, precise measurement of the empirical results. The question is how scientists can obtain such precision. The key in experiments is control over the relevant auxiliary hypotheses. Each experimental setup always depends on numerous auxiliary hypotheses, such as instruments, tools, and apparatus, but also data collection, processing, and analyses. Moreover, it also rests on various background theories and models that are simply taken for granted (Hacking 1988). Together these auxiliary hypotheses determine how precisely an experiment can measure some outcome.

Scientists can rely on a set of different strategies to ensure that the applied auxiliary hypotheses are at least approximately true. They can separate and test them by using, for instance, calibration (Franklin 1989). They can intervene in the experiment to identify artefacts (Hacking 1983). They can alter the experimental setup to build around sources of error (Galison 1987). They can repeat the experiment to better separate signal from noise. And they can vary auxiliary hypotheses, amplify them, or introduce some standards to detect false ones (Mayo 1996). Together these strategies provide a potent arsenal that allows scientists to narrow down false auxiliary hypotheses. Scientists use them extensively in practice. If the applied auxiliary hypotheses are true, so will the results from the experiment. In contrast, even if only some of the applied auxiliary hypotheses are false, scientists will receive false results.

Three factors determine the precision of an experimental measurement. First, the results from the experiment need to be true, or, from a statistical perspective, unbiased. To achieve this, the auxiliary hypotheses applied in the experiment must be true. Second, the results need to be certain. They should not vary due to minor changes in the auxiliary hypotheses. The experiments need to produce stable and robust empirical results. Major changes in auxiliary hypotheses should align with theoretically predicted changes in the outcome. Third, the results from the experiment need to be detailed. The more fine-grained the results, the more the scientists can learn. The auxiliary hypotheses should together allow for a high resolution.[2]

If scientists receive true, certain, and detailed results from their experiments, they can better achieve progress in their investigations. However, progress crucially depends also on three other important factors. They do not determine the quality of the results, but their quantity. These three factors are central determinants of all empirical testing.

First, experiments need to provide results fast. Scientists must be able to assemble or alter the network of auxiliary hypotheses quickly. If results are sparse, learning will be slow, as already argued in the previous chapter. Second, the different experimental trials need to be low-cost. The elements making up the auxiliary hypotheses need to be affordable. If an experiment is too expensive, it will not be pursued, even if it is otherwise very promising. Third, experiments need to be parallelizable. Scientists must be able to replicate and meaningfully extend each others' networks of auxiliary hypotheses. With many scientists working on some subject, this diversity greatly speeds up learning.

Importantly, these six dimensions on how results from experiments differ and how they are obtained are all necessary conditions. If in a scientific field one of them is too weak, it will result in slow or even no progress. They can compensate for each other only to some extent. For example, fast results can compensate to some degree for uncertain results. If, however, results are too noisy, learning will be difficult again.

The six dimensions differ markedly over different sciences. For example, sciences that work mostly with physical experimental setups can better separate and test, change, and exclude auxiliary hypotheses. This allows them to better narrow down false implementations of their apparatus. In contrast, sciences that work mostly with data, such as the social sciences, face more difficulties in narrowing down false choices in their data collection, processing, and analysis. Their (quasi-)experimental setups offer too many researcher degrees of freedom, which

---

[2] Note that detail and certainty in results are not the same. A result can be detailed but vary across contexts, while it can also be general but the same across contexts. However, for more detailed results to be useful, certainty needs to be high, that is, different sets of auxiliary hypotheses in the form of different studies must show similar results.

in turn cause large potential variation in the estimated empirical evidence (Simmons et al. 2011). Hence, sciences that work with physical experimental setups can produce more true, certain, and detailed empirical results than sciences that work mostly with data.

Trials in laboratories in biology do often not only deliver accurate results but are also faster, cheaper, and more parallel. Latour et al. (1979), for example, describe the work of scientists in the Salk Institute for Biological Studies. Biologists use assays for repeated trials that are not only fast but also manifold. This parallelization speeds up the entire process. Scientists treat an organism with a control substance and a new substance. The difference between the two gives numerous small-scale experiments with unbiased but uncertain results. If the substance is promising, scientists try to isolate the analogue in the brain itself through purification cycles. They then again run experiments with assays of the organism, but this time with the purified substance from the brain. Together these two types of experiments serve as mutual controls. If the recorded peaks of the original and the purified substance agree with each other, and this process can be repeated, the substance is given a name. The demand of replication to pass through both types of experiments reduces uncertainty and thus the probability of an artefact. Success is seldom though, as empirical facts are hard to establish. However, because the trials are fast, cheap, and parallel, progress is nonetheless much better achievable than in most other scientific fields, where establishing some fact that can take orders of magnitudes more time.

When they work with physical objects, scientists can discover some new result and then build on it, which in turn opens new questions that lead again to the discovery of some new result and so on, thereby always mapping their theories on the physical object. When scientists work only with theories or data, like in much of the social sciences, they cannot build on an object and discover more and more of its properties. The experimental effects are immaterial. They are average treatment effects in some population, and not something scientists can touch, that is, manipulate through trial and error. Scientists cannot observe the individual treatment effects and interact with them, build on them, or vary them to produce alternative outcomes.

More generally, sciences that work with smaller, simpler, and more clearly defined physical objects allow for much faster progress than sciences that work with larger, more complex, and less clearly defined systems. The work with small physical objects is more representative of sciences that operate in laboratories, from physics to chemistry to biology, but, importantly, also the many engineering disciplines. The work with large systems is more representative of sciences that investigate economies, societies, or politics. Whereas with small and simple physical objects empirical results arrive faster, are less expensive, and more parallel, with large and complex systems they arrive more slowly, are more costly, and remain more isolated.

In economics, for example, scientists do not have little physical toolbox economies they can run trial and error on. Most social sciences can only improve knowledge of their research objects through the long-lasting interactions between theory and data. They cannot tinker around in their investigations. This slows down, complicates, and clouds trials enormously. Sciences centered on physical objects can build up their understanding step-by-step, in a cumulative fashion. They can know where they stand, and a false route taken shows itself after some time. Sciences that obtain empirical results only infrequently receive much less guidance in their investigations. Often empirical studies do not even aim at testing some theory. Without tinkering, scientists must think through each step thoroughly in advance. For example, it takes a lot of time to setup up a field experiment, and scientists cannot just repeat it and vary it in relevant ways. Usually, scientists receive the actual results only after several years.

The sciences differ greatly in how fast they can obtain results from their trials. For example, in the computer sciences, a software algorithm might be iteratively refined with near-instant feedback on its performance. In chemistry, results can often be obtained within hours or days. In medicine, a randomized control trial might take years until some results finally become available. In modern particle physics, some experiments take decades from the initial planning until reliable results are in.

In contrast, in more practical fields like engineering science, vicarious trials have extended the reach of trial and error enormously. Engineers can rely on computers to simulate large numbers of trials. Because simulated trials are fast, cheap, and parallel, they need not even put too much effort into them. They can simply run every trial they can think of. The capabilities for ever more accurate simulated trials on computers have increased substantially over the last decades. Vicarious trials have become broader and more precise, more and more like actual trials. The trials that seem most promising can then be selected for actual trials with the physical prototype. The entire process to develop prototypes in engineering science has become much more efficient through vicarious trials. This development has sped up the advance of the practical sciences. However, the fast progress manifest in most technical sciences, visible in ever more advanced technologies, relies also on a further set of factors, to which we turn now.

6. The advantage of the technical sciences

6.1 Physical manifestation of background

In the technical sciences, physical objects in the form of technologies offer a further important way to guarantee true auxiliary hypotheses: they embody the various auxiliary hypotheses. Experiments in science commonly rely on auxiliary hypotheses that remain separate from the objects scientists want to investigate, such as instruments, tools, or apparatus, but also different ways to work with data. In contrast, when scientists construct some physical prototype, the auxiliary assumptions in the form of properties or mechanisms are part of the prototype itself. The prototype embodies them. Trials with the prototype do not impose many additional, separate auxiliary hypotheses. The physical components that make up the technology are often all there is to the trial. Scientists cannot violate physical properties and mechanisms and therefore also not violate auxiliary hypotheses. Nature sets the limits here. It does not matter whether scientists know the laws of nature or not, because they cannot not respect them.

In fact, technologies directly show whether some auxiliary hypothesis is violated, namely when the technology ceases to work or even breaks down. For example, one cannot just assume the properties of a material; either it has them or not. If they want a working prototype, scientists must respect all relevant assumptions. They must follow all theories that adequately describe nature, whether they do so explicitly or not. Nature gives in, that is, the prototype works, only when scientists manage to build it in line with those theories that align well with nature.[3]

[3] This line thought also explains why large language models (LLMs) are exceptionally good at programing. Since the web provides mostly correct code that also works, LLMs can learn the correct answers. Why is code on the web generally correct though? Because each computer program provides a separate, closed world of relatively simple laws where coding trials either work out or result in error. If these laws are violated, the code does not work. Human programmers produce code that aligns with the laws because they cannot do otherwise. They must iteratively improve the code until it works. Published,

Incorrect understanding of the laws of nature may hinder scientists in the process. Yet scientists are forced to build everything in a way that aligns with a correct understanding of nature, irrespective of whether they are aware of this or not. Knowledge of the laws of nature is thus a different matter. Some auxiliary hypotheses show themselves openly to the scientists, while others will remain unknown. Over time the understanding why things are a particular way and not otherwise will become salient. This is what makes up the knowledge of the technical sciences. Sometimes it can be expressed in rules or principles, sometimes it is more tacit. The many pieces are what constitute the craft. The close connection to a technology sets the technical sciences apart from the sciences in general.

In the case of experiments, in contrast, every trial shows results anyway, irrespective of whether they are true or not. The information about the correctness of the results rests within the entire net of auxiliary hypotheses. The violation of some assumptions therefore remains much more hidden. The auxiliary hypotheses have no ways to resist the trials in the experiment.

For example, in data analysis, the data cannot resist the estimation of a false empirical model. The data always gives in, and regression runs through, whether appropriate or not, from the very first to the very last trial. Violation of at least some auxiliary hypotheses is thereby the norm, in general without any indication of it. Most mistakes go unnoticed. The computer does not give back an error message if scientists make false data analytic choices. Scientists can thus easily misrepresent nature. Estimations can produce neatly looking statistical results irrespective of whether they are true or not. For instance, a false but highly significant coefficient in a statistically model, or also a difference-in-difference figure that looks convincing but is based on a false control group. Scientists can seldom evaluate whether econometric methods function correctly; except maybe measures of goodness of fit, but they do not inform about statistical bias of the main estimates of interest, which is the most relevant issue.

This is similar in the case of development of theories. Nothing holds scientists back from producing a false theory that does not accurately describe the laws of nature. There is no physical constraint that scientists are bound to respect, no immediate push-back from nature. Of course, the scientific method demands that a theory respects relevant empirical evidence. Ideally, it would guide scientists step-by-step toward a true theory, just like in the case of a working prototype. However, in the end scientific peers will decide about the fate of a theory. It is not just nature alone that says yes or no.

6.2 The power of “it works”

The fact that some technology “works” can generate stronger evidence in its favor than any experiment, whether in the laboratory or in the field. It provides an exceptionally strong signal to achieve coordination in the scientific community and is thus a further important reason why technology can progress so fast. It allows the respective field to gather behind and work on some more specific setup of the technology.

A newly developed technology that indeed works can end all ongoing discussions in the literature, even though ex-ante no one may have believed in the technology. We may deem a certain technology virtually impossible some years before the first tests have been implemented.

written code provides a foundation that is unambiguously true in the respective program‘s world. LLMs do therefore not also learn some false theories of coding, like with theories expressed in natural language.

But if scientists test it and it in fact works, there is no other option than to give up the skepticism. If the technology had remained just a theory, we would see endless discussions about whether it worked, with camps for and against it, and greatly differing opinions combined with severe disputes. However, when we can test whether some technology works, all scientists, irrespective of how fierce the discussions have been, must give in to the side that has predicted that it will work.

For example, Vincenti (1990) describes how in early flying machines at the beginning of the 20^{th} century the criterion for selection was: does it work? If it did, however incredible ex-ante, it established the next steps. If airplanes did not exist today, and some theory would argue that we could serially manufacture airplanes with hundreds of passengers flying at near the speed of sound, only few people would believe this. Alas, we do believe such airplanes to be possible, because we personally saw them and many of us even used them ourselves to fly. Still doubting their possibility becomes very difficult indeed.

Neural networks, and among them especially LMMs, are a technology whose capabilities would have seemed like science fiction only some decades ago. Yet they are reality now. While neural networks are not physical technologies but technologies in the form of software, their results can nonetheless always be tested for whether they work appropriately, even when mostly within the digital world itself. For instance, LLMs can produce some code that successfully executes within some specific program. We can also read the results LLMs produce and then check the facts to verify them. The industry itself tests their models by running them on a series of internal structured tasks and external public benchmarks. In fact, neutral networks are the very embodiment of constant testing in the form of trial and error. They learn by iteratively going over the training data until they can predict it accurately. The same holds for setting up the neural networks themselves. Theory is often scarce here, at least so far, such that developers improve their implementation by repeated trial and error and testing them against unused data until they work. The fact that we can directly see the power of neural networks in everyday tasks is what makes them so convincing. It becomes difficult to doubt the capabilities of the technology itself, at least in some fields like programming code, image recognition and generation, or language translation.

The power of the evidence that some technology works is a crucial difference between the technical sciences and the basic sciences. If a technology works, it is a success, and if not, it is a failure. This principle stands outside of scientific knowledge. While the laws of physics can help to design some technology, they can in no way guarantee that it also works.

Of course, there is in principle a very large number of ways in how some technology can be built. Scientists cannot follow a fixed path they just need to uncover to build a technology. Moreover, the design of a technology needs to respect constraints in terms of resources, user needs, and feasibility. Scientists therefore try out many alternative ways to construct a technology. The initial demonstration of the technology may be weak, but subsequent versions become better and better until some robust working version emerges (Arthur 2009). At the same time, different versions of the technology emerge from other scientists or companies, set up to advance it even further. Each of them will have its own working version, some a better and some a worse one. Over time, still other versions will emerge that serve more specialized purposes or markets. However, in the end, a small set of optimally working versions of the technology will dominate the market.

During the testing, the question of whether a prototype works also depends strongly on the testing environment (Collins and Pinch 2002). More difficult conditions make it harder for the prototype to count as a success. For instance, did scientists choose the most difficult scenario? There can also be different criteria for success. To obtain a convincing test for a prototype, scientists should use either a worst or best case or a broadly representative scenario.
Knowledge of why a trial worked is thereby often difficult. Scientists need to dissect and inspect the components of the technology. However, and this is the advantage of physical objects: scientists can tinker around. Often, they just run many varied trials and suddenly it works.
In contrast, whether an experiment works and produces correct results is less clear. It depends on the context, too. Yet here the context is not a distinct yes or no determined by physical operation, but an in comparison weaker answer given by, for instance, some mathematical model or empirical result. Doubting them is much easier. In contrast, what is visible by eye is very hard to ignore and dominates all other forms of evidence. The success of a technology depends only on whether it can do its intended task. It does not depend on what other engineers think of it. On the other hand, if some technology does not work, nothing can save it.
We often focus on past experiments that have worked out. This means, for instance, that they produce some phenomena in a robust way. However, these are the surviving successes from the history of science. At the frontier, scientists often have difficulties in knowing whether they have succeeded in producing a phenomenon or whether they have merely obtained an artefact. Scientists take a correct result to be one obtained with a properly working experimental apparatus. However, a properly working experimental apparatus is one that obtains correct results. Collins (1992) calls this dilemma the „experimenter‘s regress“. One depends on the other and vice versa. Scientists would need a criterion that is independent of the experiment itself. Collins (1992) claims that there are no formal criteria that one can apply to decide whether or not an apparatus is working properly. He argues that in practice the experimenter’s regress is eventually broken by negotiation with the appropriate scientific community, a process driven by factors such as career, social, and cognitive interests of the scientists.
This problem of the criterion dates back to Pyrrhonian skepticism: What do we know versus how do we know. One cannot have an answer to the first without also having an answer to the second and vice versa. Science generally embraces the second answer. It is the scientific method that guides us. In actual practice, as we have seen, scientists have many strategies at their disposal to show that their experimental apparatus functions properly. However, none of these has the same power as the fact that “it works”, which is an exceptionally strong argument. It establishes in which direction further scientific research should move.

6.3 An asymmetry in evidence

The main advantage of blind tinkering on some technology is that the scientists do not need to know whether all the involved auxiliary hypotheses are correct or not. They just try out many things, and after some time they will hit upon the right combination of auxiliary hypotheses that make the technology work in the intended way. This way scientists can construct a technology that works even without the help of much theory, that is, little understanding of why it works. To find out why it works is then the actual research task of the scientists. The fact that it works implies the existence of some theory that saves the phenomena. The physical existence of the working technology necessitates it.

How can scientists obtain an adequate theory for a technology? They could believe in a false theory that still fits somehow. However, scientists can rely on trial and error not just to construct a technology but also to create a corresponding theory. The theory develops together with the various trials. Scientists build the technology from inside out and can observe how the different parts and mechanisms map to all the involved theories and auxiliary hypotheses. The technology physically manifests them. They can also always dissect the technology in detail. If the technology is sufficiently tried out and varied, an adequate theoretical understanding will follow over time. Hence, when scientists rely on elaborate trial and error to build some technology, they often do not start out but rather end up with some theory of how the technology works. The interaction between the theory and the constant testing in turn allows for a continuous refinement and optimization of the technology.

Thus, if some technology works well in the way the scientists intended it to, and scientists have built it via sophisticated trial and error, this is evidence that the theory behind it as well as the auxiliary hypotheses are (at least approximately) true. Otherwise, the technology would not work. Some false auxiliary hypotheses may balance each other out. Yet the proper functioning along the relevant dimensions corroborates the theory and auxiliary hypotheses involved. Scientists know they did things right, and only a coincidence could have caused them to do so otherwise.

In contrast, in an experiment not focused on technology, evidence that corroborates a hypothesis can be the result of false auxiliary hypotheses, for instance some artifact that is due to external factors or malfunctioning instruments. Scientists have no objective standard they must reach, so they cannot know whether the positive evidence is a success in the sense that what they did is all correct. An assessment of one's proceeding is only possible with a benchmark that lets you know how the result must look like.

In fact, the logical asymmetry between induction and deduction, from which falsificationism profits, turns against falsification in the case of testing with many auxiliary hypotheses. If the technology does not work, the theory or some auxiliary hypotheses will be false. The problem is that we have difficulties knowing which part of the theory, or which of the many auxiliary hypotheses. The modus tollens of the falsification becomes lost in the net of auxiliary hypotheses. In contrast, if the technology works, the theory and the entire net of auxiliary hypotheses are corroborated. We do not have to search for what made the technology work. We can confidently take them all as approximately true.

Consequently, the principle that a technology "works" provides a further reason why in the development of technologies different scientists usually converge to the same set of underlying theories. It allows not only for the development of a corresponding theory that accurately describes the technology, but at the same time also for a positive evaluation of this theory.

6.4 A well-arranged puzzle

We are familiar with the idea that all scientists contribute their part to the progress of the entire field. Whereas each individual contribution may only be small, together all these small contributions form a large contribution. This is certainly the case in particular instances of practical work, where scientists or engineers contribute together to some technology that improves significantly over time. The numerous small contributions are explicitly made to fit together like in a large, complex, but well-arranged puzzle.

In contrast, in science more broadly, the small contributions of all the many scientists do not add up and complement each other so perfectly as it is the case with work on specific instances of a technology. The different contributions instead form complicated overlappings that are hard to reconcile and often even outright contradict each other. While in constructing physical objects, things are made such that they fit together, in many sciences studies are explicitly made such that they are different. For example, empirical studies in the social sciences are based on various tests of the same hypotheses, with always different research designs, data analyses, and contexts. The studies are each time based on completely different sets of auxiliary hypotheses. The numerous small contributions are not each forced by some physical structure to form an organized whole. Consequently, the different studies in a field cannot constrain each other to the same degree as work on a physical object can.

Only a series of very precise, unbiased, replicable studies will converge to a well-arranged puzzle. These will subsequently have positive external effects on each other. A series of high-quality studies might kick off a self-reinforcing process, promoting further high-quality studies. In contrast, false studies, due to for instance the choice of some false auxiliary hypotheses, will have negative external effects on each other. Finding a true effect becomes much more difficult among a set of mainly false estimates. Only a series of high-quality studies can provide the necessary mutual guidance to each other that then results in overall consistency. If the findings of different studies vary too much, the field as a whole will develop quite arbitrarily.

In the technical sciences, technologies give rise to even more technologies. They serve as building blocks for each other. Since they are forced to fit each other, progress manifests visibly in always new technologies. More theories do also give rise to more theories in the basic sciences. They also build on each other. But because they are not forced to fit each other, progress manifests itself much less visibly. One must know the entire literature in great depth to evaluate it.

7. Progress across the sciences

7.1 The increase in machinery

Repeated trial and error relies on both falsification and corroboration. Over time, a precise and coherent theory may emerge out of it. If this theory is better adapted to reality than previous attempts, it corresponds to scientific progress. Such sophisticated trial and error requires a close interaction between both theory and experiment. This ideal of a fast series of interconnected experiments was characteristic of the early natural sciences (see Chapter 4.2). Today it is still the case in some areas of biology, where theory and experiment are intricately interwoven (Weber 2004). However, in science more generally, theory and experiment have diverged more and more over the last decades. Both have become very complex and tend to follow their own respective logic. It may be a necessary development, and the only way forward, yet such a divergence between theory and experiment has inevitably slowed down scientific progress.

This divergence has been particularly prominent in the field of particle physics. For example, Galison (1987) argues that early twentieth-century experimentalists could vary their apparatus in many alternative ways. The machinery was small, inexpensive, and parallel enough to allow for extensive trial and error. In contrast, until the 1980s, apparatus had increased so much in size and cost that the ability to vary and thus rebuild it in a trial and error way has become very difficult. Earlier generations of scientists had themselves designed, constructed, and applied their apparatus. In today's large particle colliders much of it is in the hands of large teams of programmers, instrument makers, and engineers, which work together in complex relations (Galison 1997). The time scale of experimentation has increased from several months to many years since the Second World War (Galison 1987). Much of the research has thus shifted to the computer, where data analysis takes place after the experiment has already been carried out.

In modern particle physics, we thus have numerous, often very elaborate theories, but are short of the necessary experiments to test them empirically. At the bottom of the foundations of physics, facts get sparse and a manifold of theories takes over. In some areas of particle physics there has not been new, relevant data for decades (Hossenfelder 2018). Particle physicists wish they could profit from the same rapid trial and error as some fields in biology. Moreover, because of the huge size of their experiments, particle physicists shifted from exploratory research to more confirmatory research. The latter better guarantees the discovery of at least some results that justify the large investments. The detection of results is also highly automated, making it harder for experimenters to notice surprising anomalies. Together these two factors inhibit potentially valuable interaction between experiment and theory (Franklin and Perovic 2023). Things are guided top-down by theory and the room for creative trial and error is narrow, not because particle physicists choose to do so but because they cannot do otherwise.

It seems characteristic of later stage science that theory and empirics separate, with the consequence of less scientific progress. However, note that also within physics more generally, not all experimental work takes place in such huge facilities as the large particle colliders. This is specific to particle physics. Much of experimental physics still works like early twentieth century physicists, with small teams and better manageable apparatus, just like most experiments in laboratory biology.

7.2 Technological progress

The rapid progress of many new technologies we can observe almost daily is mainly due to shifts toward those technologies where trial and error is best possible. If over time trial and error in a given technology becomes more difficult, a shift toward other technologies takes place where trial and error has suddenly become possible. Progress happens where the marginal returns from trial and error are highest. The result is rapid progress in the aggregate.

Technological progress in the economy therefore never dries up because it flows like a river. It leaves areas where possibilities dry up and moves to new areas where possibilities are still greatest. This is also why large steps in technological progress always come by surprise. They seldom take place in the more mature technologies we are already familiar with; they rather create entirely new ones that were hardly even imaginable before. No one actually needs a smart phone to survive. Perfect health for example would be preferable to any kind of phone. But the former is what has been possible and, consequently, what we have got so far.

Whereas scientific progress seems to have slowed down (Park et al. 2023), technological progress is faster than ever. Much of our current scientific knowledge is already very deep, and we lack the necessary experiments to go even deeper. In contrast, because existing technologies often serve as a basis for new technologies, through combination and recombination (Arthur 2009), each new technology opens up room for even more new technologies, such as the invention of a new material that make advances in other fields possible. This becomes even more relevant with general purpose technologies such as artificial intelligence. New technology bears new technology. Knowledge of natural things does not develop in this accelerating way.

7.3 Building on physical objects

Whereas all sciences include experimentation to at least some degree, only those sciences working with small, inexpensive, and replicable physical objects will develop fast. In contrast, sciences that investigate some large, complex, and not clearly defined systems will develop slowly. The more remote a scientific field becomes from working in close interaction with well-manageable physical objects, the slower its progress will be.

The social aspect, that is, how many people are involved, is thereby not a problem as such. As long as they center around the construction of a physical object, aspects like division of labor, coordination, and management work well. The respective product or service can be properly manufactured and economic growth becomes possible. The numerous accurate results arising from the trial and error process guide organization around them. In contrast, if centered around an ill-defined system, it becomes difficult for any organization to function effectively.

Economic demand is certainly an important reason behind the rapid progress of a scientific field. However, many important scientific questions remain unanswered, even though high profits or prestige could be realized by making contributions to them. Instead, those applied science and engineering disciplines grow fastest that in fact can, namely those that build on physical objects. The richness of alternative ways to build their respective objects and the manyfold of possible applications makes them grow fast. A whole innovation ecosystem of industry and science will develop around such physical objects and not vice versa. Over time, strong science-industry relations emerge with an interwoven practice and understanding.

Hence, while a strong body of science can enable technological progress, for many engineering fields and applied sciences the causation runs from the nature of the technology to the ability to develop a strong body of science (Nelson 2008). Progress is only fast in those „fields of practice where physical artifacts - chemical substances, mechanical or electrical devices - can be made to play key roles" (Nelson 2008, p.494). Where this is the case, strong fields of applied science have emerged.

Karl Marx argued that economic growth could only take off when in the production of goods humans could be replaced by machinery. One cannot improve humans in a factory without limits, but one can do so with machinery. This made the contributions of the natural sciences, those sciences that deal with physical objects, suddenly very relevant. The core of economic growth thus lies crucially in the further development of all possible physical objects. Improve those objects that provide the most productivity gains. Continuous economic growth then follows from the boundless possibilities to always alter physical objects (Rosenberg 1974).

8. Conclusion

In the last decades, we have observed a general decline in scientific progress, although our efforts have greatly increased. The expenditures for research have risen substantially in all advanced economies. Yet we observe diminishing returns to those investments. One reason for this observation is certainly that the low hanging fruits have already been picked. Moreover, all scientific disciplines have become more specialized, such that general progress becomes more difficult. However, certain younger fields, like the computer sciences, seem to be exempt from this pattern. They still deliver impressive new results. Advancement in those fields is not held back by subject matters but rather by money. If they had more funding, they would progress even faster. This is likely not the case in many other scientific fields. This paper has argued that the main reason behind these differences lies in that fields like the computer sciences work with comparatively small, clearly defined objects in physical experimental setups, at least so far. They allow not only for results of a higher quality, but also for results that arrive in higher quantities. Moreover, these fields can profit from working with technologies, which allows for even more accurate results that also improve coordination in the relevant scientific community. The paper therefore predicts that, in every science, as soon as the objects under investigation become more remote, progress will slow down eventually. If study objects are too large, too tiny, too complex, too expensive, or too difficult to replicate, further progress becomes difficult.

Bibliography

Arthur, W. B. (2009). The nature of technology: What it is and how it evolves. Simon and Schuster.

Bird, A. (2007). What is scientific progress?. Noûs, 41(1), 64-89.

Bird, A. (2022). Knowing science. Oxford University Press.

Bud, R. (2007). Penicillin: triumph and tragedy. Oxford University Press.

Campbell, Donald T. 1974. "Evolutionary Epistemology." In: Schlipp, Paul A. (ed.), The Philosophy of Karl Popper Vol. I, 413-459. Illinois: La Salle.

Collins, H. (1992). Changing order: Replication and induction in scientific practice. University of Chicago Press.

Collins, H., & Pinch, T. (2002). The golem at large: What you should know about technology. Cambridge University Press.

Dellsén, F. (2016). Scientific progress: Knowledge versus understanding. Studies in History and Philosophy of Science Part A, 56, 72-83.

Douglas, H. (2014). Pure science and the problem of progress. Studies in History and Philosophy of Science Part A, 46, 55-63.

Fanelli, D., & Glänzel, W. (2013). Bibliometric evidence for a hierarchy of the sciences. PLoS one, 8(6), e66938.

Franklin, A. (1989). The neglect of experiment. Cambridge University Press.

Franklin, Allan and Slobodan Perovic, "Experiment in Physics", The Stanford Encyclopedia of Philosophy (Fall 2023 Edition), Edward N. Zalta & Uri Nodelman (eds.).

Galison, P. (1987). How experiments end. University of Chicago Press.

Galison, P. (1997). Image and logic: A material culture of microphysics. University of Chicago Press.

Hacking, I. (1983). Representing and intervening: Introductory topics in the philosophy of natural science. Cambridge University Press.

Hacking, I. (1988). On the stability of the laboratory sciences. The Journal of Philosophy, 85(10), 507-514.

Hossenfelder, S. (2018). Lost in math: How beauty leads physics astray. Basic Books.

Klahr, D., & Simon, H. A. (1999). Studies of scientific discovery: Complementary approaches and convergent findings. Psychological Bulletin, 125(5), 524.

Knorr Cetina, K. (1999). Epistemic cultures: How the sciences make knowledge. Harvard University Press.

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

Kulkarni, D., & Simon, H. A. (1988). The processes of scientific discovery: The strategy of experimentation. Cognitive science, 12(2), 139-175.

Lakatos, I. (1978). The methodology of scientific research programmes. Cambridge University Press.

Lamers, W. S., Boyack, K., Larivière, V., Sugimoto, C. R., van Eck, N. J., Waltman, L., & Murray, D. (2021). Meta-Research: Investigating disagreement in the scientific literature. Elife, 10, e72737.

Latour, B., Salk, J., & Woolgar, S. (1979). Laboratory life: The construction of scientific facts. Princeton University Press.

Mayo, D. G. (1996). Error and the growth of experimental knowledge. University of Chicago Press.

Mizrahi, M. (2013). What is scientific progress? Lessons from scientific practice. Journal for General Philosophy of Science, 44(2), 375-390.

Nelson, R. R. (2008). Factors affecting the power of technological paradigms. Industrial and corporate change, 17(3), 485-497.

Nelson, K., & Nelson, R. R. (2002). On the nature and evolution of human know-how. Research Policy, 31(5), 719-733.

Niiniluoto, I. (1987). Truthlikeness. Dordrecht: D. Reidel.

Niiniluoto, I. (2024). Scientific Progress. In E. Zalta & U. Nodelman (Eds.), The Stanford encyclopedia of philosophy. https://plato.stanford.edu.

Park, M., Leahey, E., & Funk, R. J. (2023). Papers and patents are becoming less disruptive over time. Nature, 613(7942), 138-144.

Popper, K. (1959). The logic of scientific discovery. Routledge, London.

Popper, K. (1963). Conjectures and refutations: The growth of scientific knowledge. Routledge, London.

Popper, K. R. (1972). Objective knowledge: An evolutionary approach. Oxford University Press.

Rosenberg, N. (1974). Karl Marx on the economic role of science. Journal of Political Economy, 82(4), 713-728.

Rosenberg, N., & Steinmueller, W. E. (2013). Engineering knowledge. Industrial and Corporate Change, 22(5), 1129-1158.

Rowbottom, D. P. (2023). Scientific progress. Cambridge University Press.

Shan, Y. (2019). A new functional approach to scientific progress. Philosophy of Science, 86(4), 739-758.

Shan, Y. (Ed.). (2023). New philosophical perspectives on scientific progress. London: Routledge.

Shiffrin, R. M., Börner, K., & Stigler, S. M. (2018). Scientific progress despite irreproducibility: A seeming paradox. Proceedings of the National Academy of Sciences, 115(11), 2632-2639.

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. Psychological Science, 22(11), 1359-1366.

Simon, H. A. (1977). Models of discovery: And other topics in the methods of science. Synthese Library, Volume 114.

Simon, H. A. (1996). The Sciences of the Artificial. MIT press.

Simonton, D. K. (2004). Psychology's status as a scientific discipline: Its empirical placement within an implicit hierarchy of the sciences. Review of General Psychology, 8(1), 59-67.

Vincenti, W. G. (1990). What engineers know and how they know it (Vol. 141). Baltimore: Johns Hopkins University Press.

Watson, J. D. (1968). The double helix: A personal account of the discovery of the structure of DNA. Weidenfeld and Nicolson.

Weber, M. (2004). Philosophy of experimental biology. Cambridge University Press.