

Proceedings of the Académie Internationale de Philosophie des Sciences

Comptes Rendus de l'Académie Internationale de Philosophie des Sciences

TOMPA

Justification, Creativity, and Discoverability in Science

Dennis Dieks

History and Philosophy of Science, Freudenthal Institute, Universiteit Utrecht, Buys Ballotgebouw (BBG), Princetonplein 5, 3584 CC Utrecht, The Netherlands E-mail: d.dieks@uu.nl, ORCID: 0000-0003-4755-3430

Abstract. According to the currently most popular version of scientific realism, the growing success of science is explained by the way successive scientific theories preserve what was true in older theories while replacing theoretical parts that have been proven false. According to this accumulative realism, it is true that scientific changes can introduce radically new ideas. But on closer inspection, there is also considerable preservation of fundamental truths or approximate truths. This view justifies the idea that successive theories get closer and closer to the truth by eliminating errors and adding to what has already been shown to be correct. Here we present an alternative to this accumulative view of scientific progress. We point out that successful parts of older theories are usually not adopted into new theoretical frameworks, but rather emerge as approximations with limited applicability. These emerging patterns are derived within a new theoretical framework that may be completely different from that of the old theory. Thus, the changes resulting from theory replacement are often more drastic than expected based on realistic intuitions. This argument casts doubt on the idea that science develops cumulatively, by accumulating more and more pieces of truth.

1 Introduction

According to scientific realism science aims at representing the world as it really is, both concerning what is observable and what is unobservable. It is an epistemically optimistic doctrine, not only saying that science has the aim of finding out the truth about the physical world but also claiming that science possesses the means to achieve this aim. Our present scientific theories, which have developed since the scientific revolution and have reached impressive predictive success can accordingly be trusted to already contain a good deal of theoretical truth. Indeed, a typical realist argument runs, it would be miraculous if science had the predictive and explanatory success it actually has, if it did not latch on to what is really going on in nature, also at the level of the unobservable. This is the so-called "no-miracles argument" for scientific realism, according to which doubting that science describes the actual mechanisms responsible for observable phenomena would amount to attributing the empirical success of science to the miraculous coincidence of finding incorrect theories that happen to yield correct predictions.

However, there is an obvious counterargument. Time and again during the history of modern science, empirically successful and seemingly unassailable theories have eventually proven to be inadequate. For example, Newton's mechanics was once considered the epitome of what could be achieved in natural science, and it seemed absurd to doubt its principled truth. It was even widely regarded as an ideal to deduce the fundamentals of other disciplines from Newtonian principles, in order to secure their truth. Nevertheless, this monument of successful physics began to falter at the end of the 19th century and has now long since been replaced by the radically different quantum mechanics. Generalizing from such cases, it appears likely that our present theories will eventually prove inadequate as well; in other words, we have to assume that they are false. This would imply that their undeniable empirical successes do not provide convincing evidence for the truth of their assumptions about underlying processes and entities. This is the so-called "pessimistic meta-induction" (Laudan 1981).

The realist camp, however, does not yield so easily. According to realists, it must be admitted that in the process of replacing a theory, some ideas about the nature of the physical world are usually overturned and some theoretical axioms are rejected; and that in this sense the replaced theory as a whole was false. But this does not mean, realists claim, that the replaced theory contained no truth. Realists claim that a detailed look at the history of science shows that not everything is thrown overboard during theoretical changes. On the contrary, some central elements of the old theory usually remain, perhaps in a refined form. Further, it should be expected that it is precisely these retained elements that were responsible for the predictive success of the old theory. Thus there is, after all, a continuous accumulation of truth. Faced with the pessimistic meta-induction, the realist needs only make a small concession, namely, that it is overly optimistic to believe in the truth of what a theory says in toto. But this does not change the fact that the success of a theory indicates that part of it is true or approximately true. One must be careful and selective and limit one's confidence to the approximate truth of those theoretical parts that were essential in producing successful predictions. These true parts are retained, which legitimizes the view that successive scientific theories get progressively closer to the truth.

This article critically examines this accumulative realist view according to which the history of science shows a continuity between successive theories that demonstrates the gradual refinement and extension of previously achieved partial truths. Certainly, we must admit that there is some kind of continuity between successive scientific theories: without it, new theories would not be able to reproduce the successes of their predecessors. However, we will argue that the continuity in question is typically the result of what is called 'emergence' in the philosophy of physics. The term 'emergence' refers to patterns and regularities that are unexpected on the basis of the fundamental laws of a theory, yet occur within a limited part of the theory's application domain; they are approximate and typically occur in coarse-

grained quantities when calculated in limiting situations. In principle, it is always possible to show that such emergent patterns lack fundamentality, in the sense that the often drastically different fundamental laws of the theory still apply and can yield more accurate predictions and explanations.

2 Retention versus emergence

(Some of the material of the following sections is also covered in (Dieks 2023c), on which the present presentation improves.)

The realist response to the pessimistic meta-induction hinges on the notion that in periods of theory change theories may well undergo drastic changes, but that a number of features such as causal mechanisms, sets of equations, or selected axioms, are typically retained and incorporated into successor theories. This preservation of theory parts is taken to indicate that the superseded theories included a kernel of truth or approximate truth. The empirical success of the older theories can be explained by their true parts (Psillos 1994, 2009, 2022; Alai 2021); that empirical success was consequently anything but miraculous, even though the older theories were strictly spoken false. As science advances, incorrect aspects of theories are gradually removed while true components are retained, extending the set of uncovered truths and improving our understanding of the world.

A standard illustration of this realist response is the transition from Maxwell's 19th-century electromagnetic theory to Einstein's 1905 electrodynamics. Maxwell's theory aimed at explaining electromagnetic phenomena as manifestations of mechanical processes, vibrations, in a material medium, the "ether", that filled the entire space of the universe. But in 1905 Einstein published his special theory of relativity, in which the same electromagnetic phenomena were accounted for without invoking any ether-like mechanical substratum. This was a revolutionary change in ontology, hard to digest for many physicists and only gradually accepted by the scientific community. However, despite this major ontological upheaval, the mathematical equations interrelating charges, currents, fields, and forces remained the same in the new theory. And of course, it was these equations that had made the successful predictions of Maxwell's theory possible; the interpretation of electric and magnetic fields as vibrations in an underlying mechanical medium played no role in the mathematical derivations. In this historical example there clearly is a theoretical core part that was retained: the relations between electromagnetic quantities represented by Maxwell's equations were left untouched. The *theoretical structure* of Maxwell's theory, defined by the relations between quantities as specified by the Maxwell equations, may thus plausibly be viewed as representing a truth already present in 19th-century electrodynamics, and as such only to be expected to survive the Einsteinian revolution (Worrall 1989). By contrast, the false assumption

that electromagnetic phenomena possess a mechanical character was rightly discarded, in agreement with the core idea of accumulative selective realism.

It should be noted, however, that the Maxwell-Einstein case is atypical: it hardly ever occurs in modern physics that portions of basic mathematical formalism remain completely intact when transitioning from one theory to another. In this respect, it is interesting to compare the following example, the transition from the 18th-century caloric theory of heat to modern thermodynamics.

The key idea of caloric theory is that heat behaves as a fluid. Heat is assumed to be a conserved substance, "caloric", consisting of very small particles that repel each other but are attracted by other matter. This theory achieved considerable empirical success (for instance, it provided elegant explanations for the expansion of materials when heated, for the fact that heat flows from hot to cold places and not from cold to hot, and for many other thermal phenomena). However, the caloric theory was completely rejected in the 19th century because its predictions failed in important cases (e.g., the production of heat by rubbing objects vigorously). According to its successor, thermodynamics, heat is not a material substance but rather a form of energy. Work, another form of energy, can be converted into heat so that heat cannot possibly be a conserved quantity. Despite this radical rejection of the core idea and ontology of caloric theory, defenders of accumulative scientific realism claim that elements of "caloric explanations" are still recognizable in explanations given by modern thermodynamics. For example, in some cases, when there is no conversion of work into heat, conservation of energy can play the same role as the earlier principle of conservation of caloric. Then again, in certain specific cases, caloric can be said to have had the same function as nitrogen in the 19th-century theory of heat; in certain other specific cases, it behaved much like modern oxygen. One might therefore argue that caloric theory was partially, approximately, and "locally" on the right track, specifying mechanisms in specific cases that bear a resemblance to what modern theory says in those same specific cases. In this way, the idea that elements of truth contained in caloric theory are preserved in successor theories may still be defended (Psillos 1994), despite the fact that the outlook of caloric theory is radically different from its modern counterparts. The case is certainly less clear than that of the Maxwell-Einstein transition however, and the claim that we are facing a case of truth retention here remains controversial (see, for example, Chang 2003, and the overview Psillos 2022, with references to criticisms contained therein; also Cordero 2011 for critical discussion of the Maxwell-Einstein case). Anyway, that successes of caloric theory can be reinterpreted by the modern theory of heat in locally structurally similar ways need not surprise us: modern theory should evidently be able to reproduce old successes, and since the mechanisms proposed by caloric theory closely follow directly observable regularities there is little reason to expect that newer theories would use structurally very different local explanations. An appeal to deeper truth seems unnecessary (see section 4).

A somewhat similar historical case may highlight implausible aspects of seeking truth in superseded theories at all costs. This example goes back to the beginnings of science. The germination of modern science is usually associated with the rejection of Aristotelianism: it is widely accepted that the Aristotelian physical world picture is fundamentally misguided and that the scientific revolution could only succeed when Aristotelian dogmas were left behind.

We will focus here on the relation between Aristotelian mechanics (Aristotle's theory of motion) and Newtonian, so-called classical, mechanics. One of the important differences between Newtonian and Aristotelian mechanics is that according to the former theory, material bodies on which no forces act persist in a state of uniform motion. Forces are therefore not needed to maintain motion; instead, they cause states of motion to change. Forces accelerate material bodies, according to the famous equation $F = m \cdot a$. By contrast, according to Aristotelian mechanics, a body will remain at rest unless a force compels it to move. Aristotle posits that forces produce a velocity, and instead of the Newtonian law of motion $F = m \cdot a$ there is the Aristotelian principle $v = \frac{F}{R}$, where v, F, and R denote the velocity of a moving body, the force exerted on it, and the resistance offered by the surrounding medium, respectively.

But even though Newton's mechanics describes the physical universe and its fundamental principles in a way that is completely incompatible with the Aristotelian view, one should expect some continuity between the two theories. Aristotle's mechanics could not have survived so long if there had been no empirical support. In fact, many everyday observations can easily be accommodated within the Aristotelian framework: objects around us do not begin to move of their own accord. We must exert a force to make them move and to maintain their motion. Empirical facts of this sort should obviously be explainable by classical mechanics as well. So, although the theoretical framework of Newtonian mechanics contradicts the Aristotelian framework, there are points of contact with regard to the description of certain patterns of events.

It is not difficult to see the details of this. In cases where a body moves through a medium that offers resistance to its motion, the Newtonian law of motion $F = m \cdot a$ must be supplemented by a friction term so that it becomes $F = m \cdot a + Rv$, where R quantifies the strength of the friction. This equation can be solved for the velocity v , and it turns out that the

solution tends toward uniform motion as time progresses.¹ If the friction is substantial, this limit of uniform motion is reached quickly; the final velocity, which remains constant, is $\frac{F}{R}$. This is exactly what the Aristotelian theory predicts. So in situations where significant friction counteracts the accelerating force, the fundamental Newtonian mechanism of force causing acceleration is obscured and it appears that force is responsible for velocity rather than acceleration. Under these special circumstances, Aristotelian relations emerge as an approximation to what is predicted by the laws of Newtonian physics.

The existence of this kind of continuity is to be expected, because Newtonian mechanics must reproduce the empirical successes of Aristotelian mechanics. Is there anything more profound to be discovered in the continuity between Aristotelian and Newtonian mechanics? Can this continuity be used to argue that Aristotelian mechanics contained a kernel of truth that Newton managed to preserve? In a trivial sense, the answer might be yes. Aristotle correctly identified certain phenomenal regularities, and these regularities were preserved by Newton's theory. This shared part could be thought of as a preserved piece of approximate truth. However, this approximate preservation of patterns is at the level of regularities in phenomena and does not represent the kind of truth preservation that scientific realists are usually after. Realism, as commonly understood, is about the discovery of basic causal factors and mechanisms in the physical world, which, accumulative realism claims, we approach ever closer through continuous and incremental improvement of our scientific theories. From this perspective, Aristotle's physics is a disaster. It fails to identify any mechanisms of motion that can be said to be retained, refined, and elaborated in classical mechanics.

3 Emergence and theory change

Emergence can be defined as the appearance of unexpected but robust patterns of behavior within certain application regimes of a theory, usually related to limiting situations of large mass, time, or length scales, or large numbers of degrees of freedom. Emergent patterns differ from the typical behavior determined by the fundamental principles of the underlying theory. Therefore, emergent behaviors, structures, or patterns need additional specific information for their explanation beyond just the principles of the given theory. This additional information may include the number of particles, temperatures, mass and length scales, boundary conditions, and the desired accuracy of the description. Coarse-grained patterns in macroscopic quantities, which differ significantly from the fine-grained, microscopic behavior primarily addressed by the underlying (sub)microscopic theory, provide numerous examples of emergent phenomena.

¹The solution is $v(t) = \frac{F}{R} + b \cdot e^{-\frac{R}{mt}}$, with b a constant and t the time.

The macroscopic gas laws are a case in point. At the macroscopic level, characterized by large numbers of particles and temperatures typical of our everyday environment, the behavior of gases is relatively simple and can be characterized by regularities in a small number of quantities (pressure, temperature, and volume). But sub-microscopically, gases are systems with many particles that generally do not behave in a simple orderly way at all.

More generally, the basic ontology of a theory, together with its fundamental laws, produces descriptions with a broad scope of application. However, emergence leads to effective descriptions that possess only approximate validity within specific and limited domains of application of the theory. The patterns that characterize these effective descriptions function as the "laws" of effective theories. From the perspective of basic theory, these are merely contingent regularities between non-fundamental and sometimes even nonexistent quantities.

Evidently, when a successful scientific theory is replaced by a new one, the new theory must be able to reproduce the successes of the first theory. For example, the successes of phenomenological thermodynamics are reproduced by statistical mechanics, and the successes of classical mechanics are reproduced by the theory of relativity and by quantum mechanics. Even the successes of Aristotelian mechanics are reproduced by classical mechanics, as we have seen. What all these cases have in common is that the old successful predictions are not exactly reproduced, but only approximated; strictly speaking, the old predictions are falsified. Moreover, from the point of view of the new theories, the old successful patterns are only conditionally valid, depending on conditions that define a narrow sub-domain of the theory's application. The old successes appear as emergent patterns, part of effective and non-fundamental descriptions.

The occurrence of emergence in the transition from one theory to the next suggests that the relationships between successive theories are usually not about refinement or incremental improvement, but involve the discovery of new conceptual frameworks not previously anticipated. Therefore, emergence challenges the accumulative realist assumption of a gradual increase of truth or approximate truth.

Even concepts that are absolutely fundamental and central in a physical theory can prove to be of mere effective and pragmatic value when the theory is replaced by a new one. A recent example of this is provided by the disappearance of the notion of an *object*, a *thing* possessing individual identity, in the transition from classical physics to quantum theory.

4 Classical particles as emergent entities

The world of classical physics, like the world of our direct experience, is a world of objects, things. Objects have definite physical properties, like position and velocity, and have definite histories by means of which they can be followed over time. In classical mechanics, the typical object is a particle—a notion that is central to the theory. No two particles can ever occupy the same position, so particles can always be told apart on the basis of where they are; moreover, each particle can be reidentified over time by means of the path it follows. Thus, classical particles, like the objects of everyday experience, are individuals.

Surprisingly, this notion of an individual object with definite properties is hard to reconcile with quantum physics.² According to relativistic quantum field theory, it is impossible to have a physical system that with certainty will be found within a spatial domain of a given finite extension (see, e.g., Halvorson and Clifton 2002, Dieks2023b). Therefore, the physical "things" that are allowed by relativistic quantum field theory cannot be localized objects. A further unexpected result is that even if we try to think of particles as non-localizable and non-classical entities, the so-called Unruh effect shows that their presence will generally be observer-dependent. For example, if an inertial observer measures a vacuum, without particles, an accelerated observer may find evidence showing that there are particles after all (Wald 1994, Ch. 5; Halvorson and Clifton 2002). This is obviously difficult to reconcile with the picture of particles as entities whose existence is objective and independent of observation.

Despite these and other seemingly bizarre results, it is clear that quantum physics should be able to make contact with the world of daily experience. The classical particle concept must become effectively applicable when transitioning from the quantum to the classical world (Dieks and Lubberdink 2020, Dieks 2023a). Indeed, there is a limiting regime of quantum theory, characterized by large masses and many environmental degrees of freedom, where typical quantum effects become difficult to detect. In this specific and limited domain, quantum mechanisms are hidden from view and the world may appear classical.

In particular, patterns in events will arise that create the impression of particle-presence. Although this happens in a very tiny corner of the total application domain of quantum mechanics, it is a corner with great significance for humans in their daily lives. But even within this classical regime, the particle picture will only work if no sophisticated experiments are performed that are able to reveal quantum effects. Quantum features remain

²In what follows we use standard interpretative ideas concerning quantum theory. There exist alternative interpretations with different roles for the notion of a particle. This situation complicates the predicament of the realist: the different interpretations are empirically equivalent, but they cannot all be true. Do some of them achieve their empirical success by some miracle? This underdetermination of theoretical structure by empirical data forms an important part of the argument against the cogency of realism, but we cannot go into this part of the argument here.

present in principle, and their detection can prove the classical particle picture incorrect.³

The situation resembles that of Aristotelian versus modern physics. As long as we do not make accurate measurements and stay within our usual everyday conditions, there seems nothing wrong with Aristotelian mechanics. But if we get precise and also look at what happens in unusual scenarios, we must conclude that reality is very different from what it seems.

5 Emergence and continuity

In the transitions from Aristotle to Newton and from classical to quantum there is certainly continuity. In both cases, old regularities are derivable from the new theory as effective descriptions, approximately valid in a small part of the new theory's domain. This may seem to confirm the continuity expectations of adherents of accumulative scientific realism, who claim that continuity is a consequence of truth preservation.

However, the example of Aristotelian mechanics as a limiting case of Newtonian theory should give us pause. There is only a small class of phenomena for which Aristotle's theory yields predictions close to the Newtonian ones. Within this domain, the emergent pattern derivable from Newton's theory is on the level of events but does not extend to mechanisms, causal links, and explanations. Aristotle's framework revolving around such concepts as natural places, natural versus forced motion, $v = \frac{F}{R}$, stands in such strong contrast to the Newtonian account that Aristotle's mechanics is often not even considered to be a part of science at all. None of the principles of motion used by Aristotle was taken over by Newton. From this perspective, the transition from Aristotle's theory of motion to classical mechanics does certainly not support the claim of truth preservation.

Nonetheless, there are phenomena within the scope of Newton's theory that can also be accommodated by Aristotelian mechanics. Doesn't this overlap cry out for explanation, and isn't the only reasonable explanation a common element of underlying truth, as suggested by the no-miracle argument? The answer is 'no'. There is an obvious alternative explanation for the continuity between Aristotle and Newton, one that does not require a shared kernel of deeper truth. This explanation is simply that Newton's theory has to reproduce the (limited) empirical success of Aristotle's theory if it were unable to do so, this would constitute a fatal objection to Newton's theory. Realists and anti-realists alike agree that successor theories must be able to reproduce the empirical success of their predecessors. This self-evident demand for the preservation of empirical success is enough to understand that successive theories must have a common part, namely the

³In fact, important progress has been made, during the last decades, in showing that seemingly macroscopic objects are actually quantum.

set of observable regularities covered by both theories. Aristotle and Newton were both able to describe bodies moving through a medium that offers resistance.

This existence of continuity on the level of observable phenomena is to be expected a priori, independent of realism or empiricism. What is more, even empiricists will expect a continuity that goes deeper than just the preservation of success at the level of the directly observable. This is because scientific theories do not contain, within their conceptual frameworks, any built-in demarcation line between descriptions that apply to what is observable by humans and descriptions of things unobservable to humans. Scientific theories have the form of objective descriptions that do not refer to observers or human perception. Therefore, it is to be expected that assertions valid for observable things and processes will also extend, at least to some extent, to proccesses and events that defy direct human observation (for example, because they are about objects that are too small to be seen). Thus, Aristotle's theory of motion predicted not only that observable heavy objects fall (striving as they are to reach their natural places) but also that invisibly small grains of heavy material will do the same. This absence of a dividing line between the observable and the unobservable applies to the conceptual frameworks of all scientific theories. Therefore, if a successor theory is able, as it must be, to reproduce the observable regularities successfully predicted by a predecessor, it should be expected to reproduce also the predictions of the old theory in a regime going beyond what is directly observable. In the example of Aristotle and Newton, the set of nearly identical predictions thus includes not only certain motions of observable bodies but also motions of unobservable objects.

Therefore, the fact that new theories are able to explain the successes of their predecessors, as emergent patterns both concerning the observable and parts of the unobservable, does not automatically imply that a piece of truth concerning the workings of nature has been preserved.

6 Emergence and truth

Accumulative realism claims that our empirically successful theories must possess a good deal of partial and approximate truth; how else could their success be explained? A considerable part of this truth comes from earlier successful theories, and these truths will be carried over again to future theories. Accordingly, we can be pretty sure that principles, processes, and entities that have withstood all theory change to date represent pieces of truth that will remain unaffected by future theoretical developments (cf. Vickers 2022). But as we have argued, there are reasons to doubt this view or at least to put it into perspective: typically, new theories transform older schemes into effective descriptions that are only approximately valid

within limited portions of the new theories' domains. Laws, principles, and mechanisms of new theories may well be radically different from the old ones. In such cases, there is no preserved truth at the level of laws, causality, and explanation. Even a basic concept like 'particle', which survived theoretical change for so many centuries, has turned out to be ephemeral.

In conclusion, accumulative scientific realism in the form we have discussed does not seem a viable account of scientific progress. The history of science shows that the empirical success of a theory may well be explainable from principles and mechanisms that are radically different from the explanatory devices offered by the theory itself.

References

- Alai M (2021) The Historical Challenge to Realism and Essential Deployment. In: Lyons TD and Vickers P (eds) Contemporary Scientific Realism: The Challenge from the History of Science, Oxford University Press, Oxford, pp 183–215, doi: 10.1093/oso/9780190946814.003.0009
- Cordero A (2011) Scientific Realism and the Divide et Impera Strategy: The Ether Saga Revisited. Philosophy of Science 78: 1120–1130, doi: 10.1086/662566
- Chang H (2003) Preservative Realism and Its Discontents: Revisiting Caloric. Philosophy of Science 70: 902–912, doi: 10.1086/377376
- Dieks D (2023a) Quantum Individuality. In: Becker Arenhart JR, Wohnrath Arroyo R (eds) Non-Reflexive Logics, Non-Individuals, and the Philosophy of Quantum Mechanics. Synthese Library 476, Springer, Cham: chapter 2, doi: 10.1007/978-3-031-31840-5 2
- Dieks D (2023b) Emergence and Identity of Quantum Particles. Philosophical Transactions A, doi: 10.1098/rsta.2022.0107
- Dieks D (2023c) Emergence, Continuity, and Scientific Realism. Global Philosophy 33:44, doi: 10.1007/s10516-023-09696-w
- Dieks D, Lubberdink A (2020) Identical Quantum Particles as Distinguishable Objects. Journal for General Philosophy of Science, doi: 10.1007/s10838-020-09510-w
- Halvorson H, Clifton R (2002) No Place for Particles in Relativistic Quantum Theories? Philosophy of Science 69: 1–28, doi: 10.1086/338939
- Laudan L (1981) A Confutation of Convergent Realism, Philosophy of Science 48: 19–49, doi: 10.1086/288975
- Psillos S (1994) A philosophical study of the transition from the caloric theory of heat to thermodynamics: Resisting the pessimistic metainduction. Studies in the History and Philosophy of Science 25: 159–190, doi: 10.1016/0039-3681(94)90026-4
- Psillos S (2009) Knowing the Structure of Nature: Essays on Realism and Explanation. Palgrave/MacMillan, London, doi: 10.1057/9780230234666
- Psillos S (2022) Realism and Theory Change in Science. In: Edward N. Zalta and Uri Nodelman (eds.), The Stanford Encyclopedia of Philosophy (Fall 2022 Edition).
- Vickers P (2022) Identifying Future-Proof Science. Oxford University Press, Oxford.
- Wald R (1994) Quantum Field Theory in Curved Spacetime and Black Hole Thermodynamics. University of Chicago Press, Chicago
- Worrall J (1989) Structural Realism: The Best of Both Worlds? Dialectica 43: 99–124, doi: 10.1111/j.1746-8361.1989.tb00933.x