Constructor Theory

Module 0

Bad Arguments Against New Science

Logan Chipkin

CONJECTURE UNIVERSITY

Bad Arguments Against New Science

Scope Is Not a Matter of Preference

With the benefit of centuries of scientific progress behind us, it is easy to miss how staggering the ambition of Newton's theory of classical mechanics must have seemed to Westerners at the time, by then only beginning to take onboard Enlightenment values. In one fell theoretical swoop, Newton had unified phenomena as apparently disparate as gravity, the dynamics of extraterrestrial objects such as other planets and the Sun, the dynamics of terrestrial objects like flying bullets and cannonballs, as well as the previously foggy concepts of momentum, force, velocity, and acceleration.

The scope of the explanation laid out in his 1687 magnum opus, *Principia Mathematica*, was not a function of Newton's arrogance, or hubris, or any other psychological trait he may have had. As with every good scientific explanation, the range of phenomena that classical mechanics explains and unifies is entirely determined by the structure of the explanation itself. One cannot arbitrarily decide which phenomena classical mechanics can, or should, account for, nor can one arbitrarily decide that it is too ambitious to be realistic.

Someone may protest that surely classical mechanics cannot account for the dynamics of a solar system different from our own by every metric, or the gravitational pull of a mindbogglingly distant star mindbogglingly more massive than the Sun. But the reach of classical mechanics is not determined by our intuitions, expectations, or feelings, but rather by the explanation itself. For example, the structure of Newton's equations do not change with increasing mass nor with increasing distance from the Earth (or any other reference point). Instead, they seem to apply to *all* objects of *any* mass.

Is it arrogant to say that the Neodarwinian synthesis applies not only to the designs we've encountered in the biosphere, but also to those we've yet to discover? To those that had existed in earlier epochs but have left no trace for us to find? To future designs? To alien life not consisting of carbon or DNA or anything that resembles earthly phenotypes? No. Not because we *want* to apply the theory to every life form in the universe, nor because we are too lazy to conjure up another theory, nor because we think that there must be a different theory for different life forms across time and space. On the contrary, it is latent in the Neodarwinian synthesis itself that the explanation for apparent design in the biosphere applies to *all* of it, past and future, terrestrial and alien. As with Newton's theory, the range of phenomena that (the modern version of) Darwin's theory explains and unifies is not up to subjective judgment but is implied by the theory itself.

A Theory of Theories

Both classical mechanics and the Neodarwinian synthesis explain, constrain, and unify regularities found in the physical world. They are theories about physical systems (or objects) with certain properties—in Newton's case, those with mass, and in Darwin's case, those that hereditarily replicate, vary, and undergo selection. But might there be regularities across the laws of Nature *themselves*? If so, could there be a theory that would explain, constrain, and unify *them*? This would not be a theory of objects, as Newton's and Darwin's are, but a theory of theories.

The principle of conservation of energy does not belong to any particular object-level theory in science. On the contrary, it constrains all of them, both those already discovered and those yet to be created. We take for granted that theories as disparate as quantum mechanics, general relativity, and electromagnetism all conform to the principle—any isolated physical system characterized by any of those theories must not create energy out of nothing, and they must possess as much energy at the beginning of their time evolution as they do at the end.

And if someone comes up with a new theory of quantum gravity that violates the principle of conservation of energy, then very few scientists will want to test the candidate theory in the lab. For the principle already serves as a kind of *theoretical* test (or criticism), one that all theories must pass to be considered viable.

The principle of conservation of energy might be the most famous of these all-encompassing constraints, but it is far from the only one. But before the advent of constructor theory, the subject of this course, our principles had been disparate and inexpressible in a single mathematical language. A theory of theories could not only unify and make precise principles we already hold dear, but it could reveal entirely new ones that could not have been thought of or expressed in the absence of such a framework.

The scope of such a theory of theories would be different in kind than that of classical mechanics yet still even farther reaching. Newton's theory covered massive objects across the entire universe, whether or not we have ever or will ever encounter them. A theory of theories, meanwhile, would cover *theories* we've discovered and those we have yet to discover (this is just a generalization of the logic of the principle of conservation of energy).

But the theory of theories could do more than 'just' help us to formalize familiar physical principles and conjecture new ones. As we will see in the next module, there are regularities in the physical world that cannot be captured by any of our object-level theories, and yet they are at least as fundamental as those that can be. The regularities that characterize information are a prime example—information can be encoded in and transferred between physical systems that are themselves described by equations from a range of theories, and yet we lack (or did lack, before constructor theory) any object-independent formalism that could capture the regularities of information itself unproblematically. It could be that the regularities of information and other such phenomena could be entirely captured by a formal framework of principles (rather than by object-level laws).

So this theory of theories would touch on not only all physical laws, but also on regularities in Nature that transcend object-level laws. Such a theory's scope is far wider than that of classical mechanics, the Neodarwinian synthesis, or indeed any theory that has come before. But, as we have seen, that is no reason to dismiss such a research program.

Opportunism Is Necessary for Scientific Progress

The basic concepts and infrastructure of thermodynamics were developed by many people over the course of the nineteenth century. By this time, Newton's classical mechanics had been well-established for generations as a pillar of what science is and should be. And yet the fathers of thermodynamics did not try to blindly copy Newton in establishing thermodynamics as a science in its own right. They did not, for example, insist on porting over the Newtonian concepts of acceleration and gravity into their fundamental laws but rather invoked concepts not found in Newton's laws, such as heat and work. Nor did they try to relate the thermodynamic quantities of heat, work, and energy via algebraic equations involving time derivatives, even though Newton had related the quantities of *his* theory via just such equations. In sum, the fathers of thermodynamics did not shackle themselves to the concepts, methods, and mathematics of the theory of an earlier age. They were rightly opportunistic, solving the problems before them with whatever explanation seemed to work, whether or not it resembled that of the great Isaac Newton.

And yet many of today's scientists, philosophers, and popularizers *do* shackle themselves in a way that the first thermodynamicists did not. For example, many assume that the next revolutionary theory in physics will be reductionistic—that is, it will explain high-level regularities in terms of low-level structure. Or they think that the next theory will consist (at least in part) of laws of motion. Or they insist that surely the next theory will have equations of one kind or another. These assumptions are so taken for granted that they often go unstated—if someone offers a new theory that is not reductionistic, does not contain laws of motion, or does not consist of algebraic equations, then it is immediately suspect.

But, absent a good explanation, we cannot know which elements of our most cherished theories will survive their successor fully intact and which will live on merely in limiting cases. For example, no amount of inspecting Newtonian mechanics could have told you that its equations of motion were merely approximations to more precise ones that Einstein would uncover in his theory of special relativity. Similarly, one could not have known that Newtonian mechanics was mistaken that quantities such as velocity, position, and angular momentum could not, as the theory implied, take on values from a continuous range but rather, as we now know from quantum mechanics, take on values from a discrete set. In both cases, it is only in light of Newton's successors that we can understand which aspects of the theory are mere approximations to a deeper truth, as well as the conditions under which the approximations work well enough for purpose.

But not all of the ideas in Newtonian mechanics were superseded by special relativity or quantum mechanics. For example, both successors held as central that dynamical equations of motion determine the state of a physical system for all time, just as Newton had. But there is no reason to think that this is itself anything more than an approximation that will not survive the next revolution in physics.

Conclusion

We have seen that the scope of a scientific theory is not a matter of subjective preference, and that judging a theory by its scope is a mistake. We have also seen that judging a scientific theory by its structure—for example, whether or not it consists of algebraic equations and dynamical laws of motion—is also a mistake. Absent a good explanation for why a given theory's scope is too great, or why a given theory's structure renders it inviable, it is irrational to dismiss a theory just because its scope or structure does not meet your preferences.

We have seen that a theory consisting of overarching principles that constrain all object-level theories, both known and unknown, and that explains phenomena that no object-level theory possibly could, is not inconceivable. Such a 'theory of theories' would have an unfamiliar and extremely wide scope, and it would have to be expressed in a language unlike one that worked for, say, classical mechanics or even quantum mechanics.

Constructor theory, the subject of this course, is precisely such a theory.

This module was just a preface, designed to preempt certain bad arguments against constructor theory—that its scope is too vast, that it does not consist of algebraic equations and dynamical laws of motion, that its structure is far too alien to be useful or true. I have not yet motivated constructor theory, nor explained some of the problems it was designed to solve.

Over the next several modules, we will examine several such problems and, for each problem, deduce the characteristics a theory should have if it is to be capable of solving it.

Thanks to Conjecture Institute Cofounder David Kedmey and Edwin de Wit for valuable feedback.