Simplicity as a Guide to Scientific Metaphysics

Jenann Ismael
jtismael@u.arizona.edu
Draft: please do not quote or circulate. Comments most welcome.

"Everything should be made as simple as possible, but not simpler." (Einstein)

"You know you've achieved perfection in design, not when you have nothing more to add, but when you have nothing more to take away." (De Saint-Exupery)

The importance of simplicity

The idea that simplicity has a role in guiding our ideas about nature is an old and recurrent theme in philosophy. Virtually everybody who has written about the construction and choice of scientific theories has recognized a role for simplicity. Editors of a recent volume on simplicity sent out surveys to 25 recent Nobel laureates in economics. Almost all replied that simplicity played a role in guiding research. [i]

Agreement on the importance of simplicity conceals a rather wide variety of different things that are meant by simplicity. Simplicity appears in a number of guises. One quite common way of thinking about simplicity is in axiomatic terms. So, it is said that we want the smallest number of axioms, or the smallest number of logically independent propositions from which a theory can be inferred. Another is in terms of the simplicity of equations. These are syntactic measures of simplicity, which is to say that they apply in the first instance to theories rather than to the physical systems they describe. One way of quantifying the complexity of a system directly would be to count its structural components, but different ways of dividing a system into structural components, and different ways of counting them up will be relevant to different classes of interactions. Another is to count the number of basic entities, or types of entity, or quantities it postulates. In recent years, these intuitive ideas have been replaced by formal measures coming out of work in computer science, information theory, statistical modeling, and complexity science. The result is a family of interconnected formal notions. So, for example, we have algorithmic information content, effective complexity, computational complexity, statistical complexity, logical depth, and others. The information-theoretic measures have introduced precision, but they are presented in terms that differ from the customary representational formats in science (they speak of length of strings in special symbolic languages where science represents with models state-spaces and equations), and so it is hard to connect them to practices in science.

There is one kind of simplicity that plays a basic role guiding the construction of models and has a special importance for philosophers interested in the ontological content of physics. I will introduce it, convey a sense of how it works, and articulate the intuitive justification for its role. I will argue that the theorist constructing a model for a complex system strives to minimize the number of degrees of freedom associated

with the system, and that the choice of physical ontology is really just a choice of *ultimate* degrees of freedom.

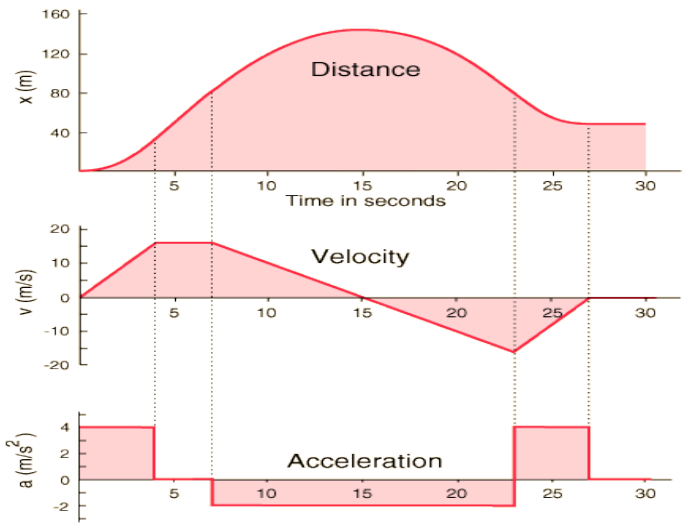
Simplicity and kinematics; model-building

The notion of a model is used in a number of different ways in philosophy of science. I will use it to mean a theoretical construction that is intended to represent the compositional structure of a system: its parts, how they are related to one another, the dimensions along which they move or vary.[ii] Examples of models include the billiard ball model of a gas, the Bohr model of the atom, the Copernican model of the universe, and the Gaussian-chain model of a polymer. Model-building is theoretical activity devoted to the exploration of different ways of reproducing the observed behavior of a complex system (or class of systems) from different accounts of its underlying structure. Model building occurs at every level of science. We develop models of signaling pathways in mitochondrial cells, the large-scale motions of tectonic plates in the Earth's lithosphere, and the digestive system of the naked mole rats. Questions about the ultimate constituents of matter – atoms or plenum? Strings or quarks? - involve model-building at the deepest level. Questions about the large-scale structure of the universe involve model-building at the grandest scale.

Let's consider a very simple example of model-building. Consider this little kinematical system.[iii] There are many different kinds of these little desktop toys. They consist of a set of weighted balls attached to rods and arranged in a configuration that produces some interesting movements.

[iv]

You can see from looking at the sculpture how complex the motions of the balls at the ends of the various rods are. In a typical theoretical context, a theorist has some information about the observable behavior of his system and he has to develop a model of the system that produces that behavior. So let's imagine that a theorist given information only about the movements of balls and he is invited to model the system that produces those movements. You can imagine these presented to him in the form of a graph. There would be one of these for each ball. Here's one that gives distance as a function of time, and then transforms it into velocity and acceleration.



or a set of functions that gives the position of each ball as a function of time.

$x(t) = \langle a,b,c \rangle$

$x_2(t) = \langle a,b,c \rangle$

$x_3(t) = \langle a,b,c \rangle$

…

If you want to imagine what the graphs would look like, just imagine the sculptures are in a black room with lights attached to each of the balls, so that what the theorist sees is 7 glowing points of light following different trajectories. The task of the model-builder is to use those movements to form an educated guess about the structure of the machine producing them. He wants to identify the independently varying components, their individual ranges of motion and reconstruct the patterns embodied in the visible movements from their mutual arrangement. He is asking how many independently moving parts does a system need to have to reproduce those patterns? What is the range of motion that each one of those parts possesses individually? What are the fixed connections among them, and how do those work together to produce the visible patterns.

Because I've displayed a picture of the machinery, we know what it looks like. But for someone just looking at the trajectories of the balls, it would be a very difficult mathematical problem to solve for the machinery that produced those trajectories. Are there three or five or ten moving parts? Is there a complicated internal engine with levers and cranks? Maybe the balls aren't connected at all? How might he go about developing a

hypothesis? The first order of business would be to get as much information as he can about the full range of motion that the system possesses. Watching the uninterrupted movements would give him some information, but that is a fraction of what he can learn by actively intervening. By intervening and observing the results of his interventions, he can come close generating movements that represent the full range of observable motions the system can exhibit. He will start the system in one configuration and then another, hold one ball fixed and see if the other ball will move, prod and probe, and push and pull. (since we want to keep him information limited to information about the movements of the balls, we'll keep the machine in the dark so he can't just flick on a light). In doing this, he is doing just what the experimental scientist does in the laboratory. We put specimens under the microscope, literally and figuratively, set up experimental conditions designed to elicit behavior that doesn't arise spontaneously, generate as much behavior in as many different settings as we can to elicit all of the discernible behaviors a system can be made to exhibit.[v] Every time he intervenes in the movements and watches the result of his intervention, he is effectively performing an experiment.

Once he has information about the results of his experiments he will start analyzing the trajectories. He will begin to notice some patterns in the form of complex conditional dependencies between the paths that the balls follow on any given run. These regularities suggest that they don't vary independently of one another. There are restrictions on their joint movements. So, for example, it doesn't matter what configuration he starts with, if ball 1 is up, ball 2 is down, and balls 3 and 4 maintain a fixed spatial distance from one another. Although in the opposite direction, the absolute value of the momenta of balls 1 and 2, and 3 and 4 is always the same. These regularities give us important clues about the underlying machinery. If we represented the trajectory of each of the balls by a line through a space each point of which corresponds to an observed location, the space associated with each ball would have 3 dimensions. Plotted together in a single space representing their joint movements, the space would have 21 dimensions, one for each assignment of position and momentum coordinates to each of the 7 balls.[vi] Trajectories through this enlarged space represent the observable motions of the machine as a whole. Each run of an experiment would start with the balls in a configuration represented by one of the points in the space and evolve through a sequence of states ending when the experimenter stopped watching or intervened to run another experiment. Restrictions on joint movements appear as points in this enlarged space that have no trajectories through them, points that the machine never visits either in its spontaneous or prompted movements. Once the experimenter has collected as much data as he can about the full range of movements his system can be made to exhibit, the model-builder is in a position to explore different ways of reproducing those movements from a categorical description of the machine itself. [vii]

How does the theorist choose between different models of the underlying machinery? Different accounts of the compositional structure of the system will come with a different account of the number of parts, the ways in which they are arranged, and the dimensions along which they vary from one another. The balls might be connected in any number of ways or not connected at all. The experimenter can never be certain that he has elicited all the movements the system can generate, and even if he had, the underlying structure is radically underdetermined by the data. So how can the model-builder form an educated guess? Here is where

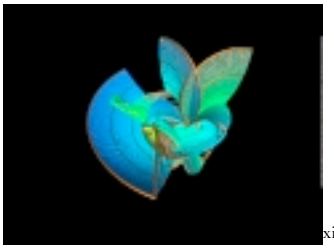
simplicity becomes relevant. The theorist looks for the most economical way of reproducing the observed motions, in a sense that is measured by of the number of degrees of freedom that he recognizes in the machine that produces them. He asks himself: How many independently moving parts, varying along how many dimensions, and with what individual range of motion do we *need* to reproduce the observed motions? How much underlying structure, measured in this way, does the observed freedom of movement *warrant* in recognizing?

The methodological maxim at work in constructing a model is the same principles engineers employ to design mechanical systems, civil engineers employ to design subway systems (well, at least in theory), and computer engineers employ to design computer programs. Do as much as you can with as little as possible. Let the need for parts be dictated by the complexity of the desired behavior. Eliminate redundancy and superfluity in your model. Excise idle wheels or superfluous parts. One might suppose that this practice rests on a presumption that Mother Nature is an elegant engineer. But that's not so. The intuitive justification is a kind of epistemic conservatism. The model builder who doesn't want to go beyond the evidence doesn't postulate parts willy-nilly. He recognizes only as many separable components as are evidenced in its observed range of motion and is agnostic about any additional structure. So far as he is concerned, Nature might be full of idle wheels or superfluous parts, but the burden of proof falls on one who wants to postulate moving parts.[viii] Every additional part that is postulated adds some freedom of movement, and if that freedom of movement hasn't been observed and can't be produced, there is no evidence for its existence.

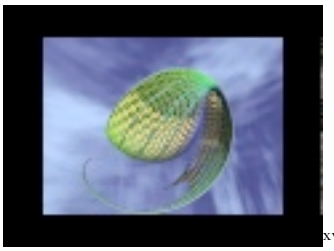
The phase space for a system is the space within which its states are represented. It has a point for each occupiable state and a dimension for each degree of freedom in that state.[ix] Even though there are 21 separate data points at any given moment to represent the visible properties of the system (one for the visible positions of each of the balls), there are only four degrees of freedom in the system itself. There is the position of the central rod with the black balls attached which moves laterally, and then the s-shaped rod, and the two appending rods to which the four colored balls are affixed. And since the movements of these parts is limited by their attachment to the base, the position of each can be given by a parameter that represents its position along what is called an orbit, which is the one-dimensional subspace of ordinary position space along which its position varies. The positions of these parts together determine the positions of the balls and the phase space for the sculpture reflects those constraints. It will have four degrees of freedom, each of which ranges over a restricted set of values.[x] Giving the values of quantities representing those degrees of freedom will suffice to determine the full state of the machine at any time, including the positions of all of the peripheral balls. The phase space of a system with four degrees of freedom will have four dimensions. And that is quite general; the phase space for a complex system has as many dimensions as there are degrees of freedom in its state. The dynamics of the machine is given in the form of rules that predict the evolution of the state from any point in the space.[xi] In a deterministic system, there will be no more than one allowable trajectory through each point. In an indeterministic system, there will be multiple trajectories passing through some points, signaling that the state of a system may evolve in any of a number of ways from some states.

If one approached the system noncommittally, reproducing the visible movements of the balls in a space that treats each of data points representing an observable property as independent degree of freedom, she would end up with a larger phase space. Since there are 7 balls, and it takes 3 parameters to specify the state of each (three for position, three for momentum), this phase space has 21 dimensions. If we plot the observed trajectories through this space we will see a great many unoccupied phase points, which is to say, points with no trajectories through them. Consider, for example points in which ball 1 and two are contiguous or points at which balls 3 and 4 are both at the highest points of their trajectory.[xii] These represent states that the system never occupies, and they are the model-builder's clue to the compositional structure of the system. They expose constraints on joint values of what are being treated as independent degrees of freedom. Moving from the space in which we simply represent the observable properties of the system to the space in which we represent its real underlying state involves a reduction in the dimensionality of the phase space and concomitant excision of empty space. A theorist can be sure that he has captured the real degrees of freedom in a system only when he can reproduce the full set of observed trajectories without redundancy (observationally indistinguishable motions are not represented by distinct phase space trajectories) or loss (observationally distinguishable motions are represented by distinct phase space trajectories) from a phase space with no unoccupied points. That is an ideal seldom realized, but it is an ideal at which a good deal of theoretical practice is aimed. [xiii] Cognoscenti will recognize that both of Einstein's relativity principles were intended to expose and excise this kind of redundancy.

In kinematical contexts, degrees of freedom are moving parts and dynamical behavior is motion. In non-kinematical contexts, degrees of freedom are represented by adjustable parameters that can also represent internal quantities and relations among the components. Kinetic sculptures give us one way of seeing how a machine with a relatively small number of independently varying components could produce apparently complex visible movements and how a theorist trying to develop a model uses regularities in those movements as clues. But equally striking non-kinematic examples are available from the virtual world. Computer generated desktop screen savers are produced by simple programs in which values of a small number of parameters generate complex patterns of changing color across a pixilated screen:

 [xiv]

Or

 [xv]

We could reproduce the patterns generated by a given screensaver in a space treating the color of each pixel as an independent degree of freedom. Such a space would have at least as many dimensions as the number of pixels.[xvi] But it would also have a good deal of empty space, i.e., unoccupied points representing screen configurations that never arise. The theorist trying to model the program that generates the patterns will look for one that visits all and (as close as possible to) only the regions of the space of all describable assignments of colors to pixels *that are actually observed.*[xvii] There is no *a priori* way of doing this. He needs to observe the dynamics, to see which configurations arise and how they evolve. Only a careful mathematical analysis of the actual trajectories will tell him how many independent degrees of freedom there must be in the program that produced them. If he recognizes too few he will not be able to reproduce all of the trajectories. If he recognizes too many, he will be recognizing freedom of movement for which he has no evidence. He could, of course, postulate laws that restrict the relative motions of components, but if he can exclude those trajectories without the need for laws that place ineliminable restrictions on separate degrees of freedom, that is far and wide the preferable option. (I'll say something more later about why this is so, but the intuition is that the need for fundamental restrictions of this kind reveals that you haven't really captured the degrees of freedom. It is a reason to keep looking for a simpler model).

From phenomenological data to data models

Let's describe the procedure abstractly in general terms. The native position of the theorist is to be in possession of data about a system's dispositions to produce observable behavior.[xviii],[xix] From the bundle of loosely correlated dispositions, scientists generate 'data models'.[xx] A data model is a corrected, rectified, regimented, and in many instances idealized version of the data gathered in the lab and out in the wild. In a data model, messy, complex phenomenological data is transformed into something clean, relatively compact, and expressed in mathematical terms.[xxi] Information is separated from noise, errors are removed, and curves are smoothed out. Measurements are made carefully and precisely, checked and systematized. As above, data models can be expressed in the form of graphs or equations. Equations that give the values of a collection g of measurable parameters as a function of time can be graphed in a g-dimensional space. Once the theorist has a full, clean, compact representation of the observable dynamical behavior of his system, he is in a position to analyze the trajectories and come up with a reduced space that captures the true degrees of freedom in the system that produces the data. I'm going to borrow the term Beable from John Bell, who introduced it into discussions of quantum mechanics to distinguish quantities that represent the real categorical properties of a system from its dispositions to affect measuring instruments in certain ways.

The task for the modeler is a generalization of the process we saw above with the kinetic sculptures. He needs to reproduce the regularities implicit in data models in a relatively low-dimensional space defined by the smallest number of independent Beables.[xxii] Insodoing, he is navigating between two constraints; the need to recognize enough structure to ground the observable behavior and the prescription not to recognize more structure than is strictly needed to do so. He wants to end up with a phase space defined by Beables rich enough to produce all observed trajectories through his data models (and hence rich enough to ground all measurable differences in observed behavior), but he will cut away as much representational fat as he can. This is what Einstein was pointing to when he wrote:

"[T]he supreme goal of all theory is to make the irreducible basic elements as simple and as few as possible without having to surrender the adequate representation of a single datum of experience."[xxiii]

There is no guarantee that this procedure will be rewarded. If every observable varies independently of any other, there will be no way of reproducing the data from a reduced set of Beables, but any regularity, no matter how weak, i.e., any restriction on joint values of observables, presents potential to simplify.

Ultimate vs. constrained degrees of freedom: simplicity and fundamental ontology

When we are dealing with a local subsystem in a fixed setting and considering a restricted set of interventions, the range of motion is often limited by the fixed connections among the parts. But 'fixed' is a term that is relativized to a class of interactions. What is fixed relative to one set of interactions is not fixed relative to another. In the example of our kinetic sculpture, we came up with only four separable components because we considered only interactions that left the integrity of the sculpture intact. We imagined gentle interventions that imparted motion to the balls, but nothing that damaged the fixed connections among the parts, broke the rods, or crushed the balls, and we said nothing of magnetic fields, extreme temperatures or gravitational forces. When we widen our consideration to consider the full class of possible interactions that we could perform, balls will come off the rods, will themselves be decomposed into molecules, electrons, and perhaps ultimately quarks or strings. New degrees of freedom may have to be recognized to explain magnetic gravitational and thermal phenomena. It is only when we widen our gaze to consider the full class of physical interactions that we isolate the ultimate degrees of freedom in nature.

The day-to day-life of most practicing scientists is focused on a particular system or class of systems. The special sciences, for example, model systems at a particular scale, usually under restricted conditions. They aim for a detailed and specific account of a proscribed range of phenomena relative to a set of constraints, and don't consider interactions that disturb those constraints. The constraints can be tacit or sometimes built loosely into the description of the system. They are not hard and fast, but they do guide the focus of research. The biologist studying cells in vitro, for example, is interested in the behavior of cells in a very specific range of settings. She considers interactions that keep the integrity of the cell intact and alive. Once she moves beyond those interactions, she is no longer studying the cell *in vitro*. The crystalographer studying

quartz will consider interactions that don't disturb the lattice structure of his materials. Once he moves beyond those interactions, he is no longer studying quartz.

When we add the models of the special sciences together, each with its own scale and constraints, we have a motley and quite complex set of partially overlapping models. Per Bak remarks on the remarkable variety the world displays from one place to another, and at different levels of decomposition. This variety is captured in local submodels that study the dynamics of these structures relative to different sets of constraints:

"The surface of the earth is an intricate conglomerate of mountains, oceans, islands, rivers, volcanoes, glaciers, and earthquake faults, each of which has its own characteristic dynamics. Unlike very ordered or disordered systems, landscapes differ from place to place and from time to time. … If we zoom in closer and closer, or look out further and further, we find variability at each level of magnification, with more and more new details appearing."[xxiv]

Constraints go along with classes of interactions. For every constraint, there is an interaction that disturbs it. And it is notable that this de facto variety is maintained only because (and so long as) these systems exist in environments in which they are shielded from interactions that break those constraints. Mountains, islands, rivers and glaciers would not stay mountains, islands and rivers and glaciers if powerful enough bombs were dropped, the temperature got high enough or gravity was considerably weakened or strengthened.

We distinguish degrees of freedom ***relative to constraints*** from ***ultimate*** degrees of freedom. The former are degrees of freedom in the state-dependent properties of a system so long as we consider a restricted class of interactions. The latter are the degrees of freedom that there are relative to the full class of physically possible interactions. The only constraints that remain when we consider the full class of physical interactions are those defined by fundamental law. The degrees of freedom associated with these local submodels are relative to constraints. But the models are unified at the deepest level by fundamental theories in physics. Where special sciences aim for specificity, physics aims for generality. In a fundamental theory, all constraints are dropped. When we are studying a locally constrained subsystem, we are asking: 'how much freedom of motion a system exhibit under fixed constraints?', 'what are the degrees of freedom the system possesses relative to interactions that preserve those constraints?'. In a fundamental theory, we are asking 'how much freedom of motion does nature possess when we remove all constraints and allow the full class of interactions?', 'what are the *ultimately* separable components of matter? what internal degrees of freedom do they possess? and how much freedom is there in their arrangements?'. And our answers have to tell us how those components have to be arranged to give rise to the locally constrained subsystems encountered situationally and studied by the special sciences. It is at the fundamental level that we get an answer to Bak's question: "How can the universe start with a few types of elementary particles at the big bang, and end up with life, history, economics, and literature? … "[xxv].

The procedure by which the theorist arrives at an answer at this level is not different than the one that is used at other levels. She has the task of trying to reproduce the full variety of dynamical behaviors captured in the constrained models from a reduced set of underlying Beables. This is a much more general setting than we saw with the kinetic sculpture. The scope is unrestricted. Although there is more data in terms of volume, it is also thin relative to the scope of the theory. There are all kinds of interactions that a fundamental theory covers that don't occur naturally and are hard to recreate experimentally. Epistemologically, we arrive at a fundamental theory by working simultaneously downwards by decomposition and upwards by reconstruction. The only empirical constraint on fundamental theory is that it generate the constrained, higher-level models that make the most direct connection with phenomena. The degrees of freedom associated with a fundamental theory are provided by its ontology. This will identify the ultimately separable components of matter with their internal degrees of freedom and delimit the range of configurations in the form of a catalogue of basic entities, quantities and relations.[xxvi] Criteria for choice of fundamental ontology are no different from the criteria we saw above in constructing models of local subsystems, and simplicity plays the same role. We do everything we can to generate the full range of motion, now allowing unrestricted kinds of intervention and exploring new possibilities for recombination. The phase spaces of subsystems studied in the special sciences, and the kinds of systems encountered situationally like kinetic sculptures, cars and can openers are reconstructed as subspaces defined by particular constraints. We are not just interested in configurations that arise naturally; we ask what are the possibilities for recombination into new systems? What sorts of macroscopic systems would we expect to see allowing unrestricted recombination of ultimate parts, according to the theory?[xxvii] And we try to produce those possibilities.

If the ontology that is offered at the fundamental level - the basic entities, quantities and relations - really does isolate the basic degrees of freedom, there should be no *ultimate* restrictions on their recombination. We ask ourselves what variety of phenomena would we expect to see from arbitrary recombination of basic entities, each exercising its full range of motion? Any misfit between expectations derived from unrestricted motion and what is actually observed demands explanation. There is a variety of more or less 'accidental' reasons that a possible configuration or state doesn't actually arise. But if we can't produce the configuration artificially and can find no explanation for the failure to observe or produce it, the suspicion will be raised that the theory has some excess representational capacity. And if a simpler theory is proposed that reproduces the observed behaviors without the unused representational capacity, that will be a mark in its favor. The intuitive justification here is that any restriction on combinations of values for what are being treated as independent Beables constitutes a violation of the Humean ban on necessary connections between distinct existences. And whatever can be said about that ban in other contexts, in this context, it functions as our best guide to identity.

The goal of unrestricted recombination in composition and full exploitation of the phase space of a fundamental theory acts as a regulative ideal, never fully realized. But it is the rationale for the insistence that if a theory allows the existence of something, we ought to find it in nature. In 1931 Paul Dirac showed

that the existence of monopoles is consistent with quantum mechanics, and argued that if their existence is not ruled out by theory and that "under these circumstances one would be surprised if Nature had made no use of it".[xxviii] This sort of reasoning is widely—though not universally—accepted by other physicists. One often hears claims like the following:

"One of the elementary rules of nature is that, in the absence of laws prohibiting an event or phenomenon it is bound to occur with some degree of probability. To put it simply and crudely: anything that *can* happen *does* happen. Hence physicists must assume that the magnetic monopole exists unless they can find a law barring its existence,"[xxix]

The justification for such claims is the same principle of selection at work in the choice of fundamental ontology that we saw in the preference for the simpler model of the kinetic sculpture. Any selection is defeasible, pending data that requires the recognition of new degrees of freedom. The experimentalist is always trying to push the boundaries, performing experiments under increasingly extreme conditions both to see if he can realize unobserved phenomena whose physical possibility a theory predicts and phenomena that might disprove existing theory. The theorist is always trying to refine foundations, come to a clean and unambiguous understanding of the underlying ontology that recognizes only as many degrees of freedom in nature as we have evidence of. It acts in concert with the need to accommodate the phenomena that have been observed.

This relationship between the fundamental theory and the theory of local subsystems is expressed formally in Lagrangian mechanics. Lagrangian mechanics is a re-formulation of classical mechanics that combines conservation of momentum with conservation of energy. In Lagrangian mechanics, the trajectory of a system of particles is derived by solving the Lagrange equations, which incorporate the constraints directly by judicious choice of generalized coordinates.[xxx] In the study of multi-body systems, generalized coordinates are a set of coordinates used to describe the configuration of a system in relative to some reference configuration. For a set of coordinates to serve as generalized coordinates they must *uniquely* define any possible configuration of the system relative to the reference configuration. The number of independent generalized coordinates gives the number of degrees of freedom in the system relative to the reference configuration. [xxxi] The physics of the system is independent of the choice of generalized coordinates, so we choose coordinates that make the equations of motion easy to solve for a given problem. What the generalized coordinates do when we are dealing with a subsystem relative to a set of constraints is effectively allow us to keep track of only dynamical variables make a difference to evolution. In technical terms, they allow us to ignore constants of the motion. Consider, for example, a small frictionless bead traveling in a groove. If one were tracking the bead as a particle, calculation of the motion of the bead using Newtonian mechanics would require solving for the time-varying constraint force required to keep the bead in the groove. For the same problem using Lagrangian mechanics, one looks at the path of the groove and chooses a set of *independent* generalized coordinates that completely characterize the possible motion of the bead. This choice eliminates the need for the constraint force to enter into the resultant system of equations. There are

fewer equations since one is not directly calculating the influence of the groove on the bead at a given moment.

When we go from the generalized coordinates of a constrained local subsystem of the world to the equations of a fundamental theory, we think of ourselves as releasing constraints that held ultimately separable components in the relatively fixed configuration that defined the system. From the perspective of the Newtonian equations, the generalized coordinates are the form the equations take relative to constraints. The fundamental laws are what we get when we remove all constraints. The number of degrees of freedom that we have to ultimately recognize as physically present to reproduce the dynamics of all of the locally constrained systems that we observe are what the fundamental equations give us. Any apparent restrictions that remain in that setting – e.g., combinations of values for basic parameters that are never observed in conjunction, or configurations of basic entities that never arise and can't be produced - represent necessary connections between distinct existences and they are regarded with suspicion by the theorist. If we can produce a simpler theory that reproduces the full range of observed motion without generating unobserved excess, then the simpler theory is doing a better job of capturing the structure that we are seeing manifested in the phenomena. Because of the complexity and holistic nature of physical explanation, there may be no obviously idle wheels in the theoretical machinery, or no simple way of expressing the difference between what is allowed and what is observed (e.g., no compact expression that says these values never occur in conjunction or these configurations never arise) but the mere existence of a way of reproducing the full range of observed phenomena with fewer degrees of freedom, suggests that there is some unwarranted complexity in the machinery. And that motivates preference for the simpler theory.[xxxii] There is no way to know whether one has arrived at the simplest theory.[xxxiii]There is only a negative test. The only way to know you haven't gotten the simplest theory is to produce a simpler one. And even then the choice between them is not straightforward. Choice favors the simpler theory only *ceteris paribus*, and in practice, unless theories are empirically equivalent, *ceteris* is never *paribus*. Even when they are empirically equivalent, they may have different potentials for development. Theories are undergo constant testing and revision; new predictions are derived, new tests constructed, and accommodations made. No theory is empirically perfect and one must weigh simplicity in together with comparisons of empirical fitness.[xxxiv] But it still stands out as an important guide in model-building.

Ontology led by dynamics; this is not yo' mama's metaphysics.

Pre-scientific metaphysics was led by imagination. For Thales, Aristotle, Heraclitus, and Abelard, imagination was the primary tool in trying to develop insight into the nature of the world. Newton was the first to exploit the power of mathematics with great success. Whitehead said of Newton that what was most characteristic of him as a scientist was that he had a nose attuned to the mathematicizable substructure of the world. Nowadays, mathematical methods are indispensible as a guide to the theorist. The imagination is only so good at inventing *ab nihilo* new ways of representing things. As complexity increases, visualizability decreases. Mathematics is needed both to provide compact ways of representing the vast bodies of data we

now bring to bear on theorizing, and new ways of recognizing and disclosing regularities in the data. I have been emphasizing that those regularities aren't simply used in the formulation of laws. They are guides to ontology. They help us minimize the number of degrees of freedom that we recognize in the system. Where the imagination starts with an ontology often derived by extending common-sense ideas to the microscopic world and defines a dynamics by introducing rules restricting behavior, physics gives us techniques for starting with the dynamics and developing an ontology is fitted to the dynamics. The goal in choice of ontology is that much of the regularity in the behavior emerges without the need for laws. The abstract dynamical analysis that I described above *starts* with the trajectories and builds a phase space that strives to eliminate phase-points that would otherwise have to be ruled out by law. We start with a phase space that has a separate degree of freedom for every observable quantity. We get as close as we can to a full set of trajectories for the system plotted in this space. And then we look for ways of reproducing those trajectories from a reduced space by whittling away the excess structure.

Just as there is no way to know whether we have the simplest theory, there is no finite effective procedure for producing a simpler one. But there are quite powerful, specific techniques for recognizing dynamically irrelevant structure. Einstein spoke often of simplicity as a kind of 'inner perfection' in a theory. [xxxv] In his own theorizing, he strove to weed out degrees of freedom not needed to reproduce the phenomena, pioneering the most powerful methods of recognizing and removing unnecessary complexity.[xxxvi] The use of symmetries to identify and remove structure that is not supporting distinguishable dynamical behavior has been exploited with particular power and studied in the context of space-time theories.[xxxvii] String theorists and particle physicists have exploited the same techniques and added a whole arsenal of new ones.[xxxviii] There is a lively and ongoing debate in the philosophical literature in the foundations of physics about the conditions under which these methods are appropriate. But when they can be implemented without jeopardizing empirical adequacy, these kinds of formal methods play an increasingly important role guiding us towards simpler formalisms.[xxxix]

All of these are ways in which the metaphysics is guided by the dynamics. Modern physics has departed in quite radical ways from the ordinary ontology of everyday things, led by the kinds of criteria that I'm talking about here. In relativity the basic entities are no longer three-dimensional objects, but events or perhaps space-time points. Time and space are unified into a single-four-dimensional manifold in which temporal and spatial distances were replaced by spatiotemporal distances and forces are incorporated into the geometry of space-time. Proposals coming from quantum gravity depart even more radically from common sense ideas. In modern physics, common sense plays almost no role in choice of basic ontology. The physicist pays attention to what the dynamics is telling him. Some intriguing new kinds of hypothesis get rid of the space-time framework altogether except as an emergent structure. The dynamics is defined over an ontology of particulars that are not localized in ordinary three-space. One need not, of course, acquiesce in this approach to ontology. I am exhibiting these methods because seeing physics as guided by the attempt to reconstruct observed trajectories in a manner that is guided by an ideal of simplicity gives some insight into how ontological reasoning works in science.

Identity and individuation

The freedom of movement in a complex system is determined by the number of parts, the dimensions along with they vary, and their range of motion along each dimension. These together place constraints on the range of motion of the system as a whole. To the model-builder approaching a system of unknown structure and composition, it is an open question how many separable components a system has, the dimensions along which they vary, and their range of motion along each dimension.[xl] Each separable component, and each of the quantities that characterize them internally is an independent Beable. Beables have dispersed and wide-ranging effects on surface phenomena, always in combination with other beables. The scientist has the job of disentangling their effects, identifying the basic components, the dimensions of freedom in their internal states, and the freedom they exhibit in configuration, and reconstructing the surface phenomena from their joint activity.

The methodological principle that the theorist applies is: minimize coincidence and avoid fundamental laws that place ineliminable restrictions on separate degrees of freedom. This turns into a prescription to explain every regularity from laws of evolution defined for the fundamental entities, assume no ineliminable laws of coexistence. The justification has to do with identity and individuation. The theorist is trying to sort out the underlying structure but her grasp of Beables is mediated by their surface effects. She has no guide as to the identity or distinctness of Beables responsible for different effects other than the relations of covariation among them. She does her best to isolate the effects of particular Beables experimentally, constructing measuring instruments and setting up laboratory situations that are intended to track them individually. But it is a messy business. Any attempt to isolate the effects of one Beable has to shield from the effects of other Beables, and there is no theory-neutral way to be sure that one is seeing the effect of a single variable. Experimentalists are in the position of fishermen, casting lines below the surface of a murky pond, trying to get their hooks into fish whose movements are producing ripples on the surface. There's nothing to tell them beforehand, how many fish there are and how freely they move. How many of the lines they have below the surface have got their hooks into the same fish. In principle if two lines are always pulled at the same time and in the same direction, one would think that they are tracking the same fish, but lines cross and interfere with one another. One has to sort out which of the movements are due to the movements of the fish at the end, and which are due to interference from other lines. The theorist has to effectively solve simultaneous equations for all of these things. And although there is no guarantee of success, a combination of experimental ingenuity and theoretical insight guided by a preference for simpler models is one way to approach it.

The Humean ban on necessary connections between distinct existences plays an important role in this process not because it is an a priori truth.[xli] we are not committed to denying that there couldn't *be* entities that bore one another necessary connections. It is that it functions in this context as the best test we have for identity.[xlii] Independent variation becomes a criterion of individuation for Beables because access to Beables is mediated by their effects. When we see fundamental restrictions on recombination, we suspect that our

principles of individuation are not lining up as they should with the evidence. We have *empirical warrant* for recognizing distinctness where we see evidence of independent variation. Beyond that, we remain agnostic.

This is a very abstract way of getting a handle on underlying structure. It's something that our brains do quite naturally and un-self-consciously. We look for patterns in the phenomena not immediately as evidence of laws, but first as clues to underlying identities. The police detective noticing correlation in the evidence at different crime scenes; similar crimes, same handprints, tire tracks, blood type, and shoe size. His first suspicion is that the same criminal is responsible for both the crimes. A teacher hears multiple reports of a brown dog in the hallway, until she has evidence of observable difference her first assumption is that they concern a single animal. The doctor noticing symptoms of disease assumes, until forced to do otherwise, that they have a single underlying source.

Explanation

Regularities can take the form of correlations, covariation, and conditional dependencies among the values of Observables. Any apparent pattern or restriction on the joint values of Observables is a form of regularity. The physicist looks for explanation of regularities.[xliii] There are two ways to explain a regularity; one is to derive them from laws that restrict coevolution of independent Beables. The other is to have them emerge from unrestricted coevolution of a reduced number of Beables. In the first case, I will say they are explained by law. In the second case, I will say they are explained *organically*.[xliv] All of the examples I pointed to above (the police detective and the crime scenes, the dog in the hallway, the doctor and the disease) are organic explanations. Explanation by law traces regularities to connections between distinct existences, organic explanation traces them to underlying identities or what we might call 'constitutional necessities'. Constitutional necessity is a generalization of identity that recognizes overlap.[xlv] So, for example, the fact that all Parisian inhabitants are also inhabitants of France or that all red objects are also colored, follows from relations of constitutional identity between Paris and France or red and color. A theory that explains all observed regularity organically derives conditional dependencies embodied in distributed patterns in the manifold from underlying constitutional necessities.[xlvi]

The result of targeting simplicity is a model in which organic explanation is optimized. In the limit, all regularities in surface phenomena should be emergent from laws of evolution applied to distinct existences. This is a way of saying that constraints on the 'range of motion' of the whole should emerge from laws that describe the evolution of components with no additional restrictions on joint evolution.[xlvii] A theory that recognizes fewer underlying degrees of freedom will explain organically some of the regularity that a more complex theory will either not explain, or explain by laws that recognize necessary connections between distinct existences. The more underlying degrees of freedom we recognize, the wider the range of trajectories produced by unrestricted coevolution, and so the more restrictions have to be placed by the laws to explain the regularities. The best type of theory is one in which all regularities without the need for laws that impose restrictions on independent variation at the fundamental level.

The preference for organic explanation guides us in mundane examples of inference to the best explanation. If you see regularities assume don't assume they are coincidences or coordinated action between distinct existences. Start with the assumption that you are seeing redundant manifestations of the same underlying degrees of freedom. See if you can trace them to constitutional necessities. A detective forming a theory of a crime identifies the shot heard by one witness with the flash of light seen by another and the gunshot whose effects are witnessed in the aftermath. Of course it is possible that they are records of three different events, but if they can be explained at once by a single shot, that is the preferred explanation. A team of detectives comparing notes, noticing correlations across a number of crime scenes (exactly the same *modus operandi*, same shoe size, hair color, tire tracks) will take the evidence to favor the identity of the subject. In all of these cases, the hypothesis that explains regularity organically by tracing them to underlying identities is preferred. The preference is defeasible, always pending further data, or some extrinsic reason for recognizing distinctions. It often happens in science that when we think that we have things nicely tied up, nature reveals unsuspected freedom of movement.[xlviii]Any presumption of identity relative to a given body of evidence can be defeated by new evidence.

One very beautiful example of organic explanation is the explanation of the regularities among of sensory streams. One has to abstract from one's native realism to appreciate the complex web of conditional regularities among visual, auditory, tactual, kinaesthetic experience. You know instinctively when you see an object in front of you that if you move in the right way (walk towards it and put out your hand), you can expect a certain tactual experience. You know that if you turn your head, keeping your body and environment fixed, your visual experience will change in certain ways. When sensory streams are derived from a model of the material world, the complex, conditional regularities among them emerge organically from the way the body is put together and related to the environment. Likewise for the explanation of the regularities in experience among separate and differently located subjects. Those regularities are quite complex and invariably conditional. You and I don't have the same experience, but if we are located in the same vicinity, looking at the same landscape from different locations, there will be a complex network of relations that emerge organically from the way we relate to our common environment when embedded a single model of a common world.[xlix]

Preference for simpler theories unifies other explanatory principles

A number of methodological principles and practices in science can be unified by seeing them as forms of organic explanation. Many of the intuitions that guide everyday plausibility judgments are properly characterized in these terms. Explanations seem *ad hoc*, needlessly complex, or baroque explanation if they incorporate needless complexity. I also mentioned Principles of Plenitude. There should be no fundamental restrictions on recombination and no fundamental restrictions on occupiable states. Every real degree of freedom recognized in the state of the system and every real combinatorial possibility for constructing new systems should be exercised. Principles of quantitative parsimony (don't postulate more basic entities or independent components in a system than are needed to explain its behavior), and qualitative parsimony

(don't recognize more ultimate types in nature than are needed to reproduce the full variety of observed phenomena) can be seen as auxiliary to the preference for theories that recognize the smaller number of ultimate degrees of freedom.

I haven't talked about causal explanation because the distinctions I've been drawing are prior to any talk of causation. In everyday contexts, we explain individual events by citing causes. When we ask what caused the power to fail or the driver to swerve, we want to know what brought it about, and we cite earlier events that bear it the right kind of historical relation. But at the fundamental level in physics, causal relations do not appear. If physics talks about causes at all, it talks about them only implicitly as a special class of emergent regularity. It is hard to say anything general about the nature of causal relations because the issue is so hotly contested, but for our purposes, there are two broad schools of thought. One is counterfactual. The causal effects of a physical parameter A are downstream events that bear a special sort of counterfactual supporting relation to variation in A.[l],[li]. The other way is to thinking of causation as a special kind of *process*. Again, the details can vary, but a causal process is typically regarded as a temporally asymmetric chain of contiguous events, each of which is produced by the preceding. One of the most familiar maxims of everyday causal explanation is a second-order principle that concerns how the causal explanations of several different events should relate to one another. It says that when a pair of events (E1 and E2) are correlated, the *correlation* should be explained by the by finding a common ancestor, an element, C, in their causal history, conditionalizing on which renders them probabilistically independent. So, for example, if 40 people within a 20 mile radius come down with food poisoning, even though there could be separate causes for each illness, separate causes for each illness leaves us with an unexplained correlation (between living in the radius and getting sick). Identifying the causes explains the regularity organically without any need for coincidence or additional laws. This intuitive maxim of everyday reasoning plays an important role in science and was canonized by Reichenbach as one of the central rules of scientific inference. He argued that every pair of correlated joint effects must be explained by a screening off by a common cause.[lii] Justifications for the principle have been discussed quite heavily by philosophers of science. For our purposes, we can note that common cause explanations are a form of organic explanation. When we explain a regularity from an underlying identity, we reduce the number of degrees of freedom that are needed to explain the effects explained.

Another theoretical principle that can be seen as a preference for organic explanation has to do with a prejudice against a certain type of law at the fundamental level. There are two types of laws: *Laws of Succession* which relate successive states of a system and can be thought of as regularities in evolution, and *Laws of Coexistence* which restrict the values of quantities at the same time. Examples of laws of succession are Fundamental Laws of Temporal Evolution (Maudlin's FLOTE's).[liii] Examples of Laws of Succession are Gauss' law relating the electric flux f through a closed surface generated by an electric charge q located within the surface ($f=e_0q$) and Boyle's law relating pressure (P), volume (V) and temperature (T) ($PV=kT$, with $k={}^{1}/_{2.7315}$). Laws of Coexistence are tolerated in a phenomenological theory like thermodynamics, but there is a strong prejudice that they should be eliminated at the fundamental level. In a fundamental theory,

all laws should be Laws of Succession (preferably having the form of differential equations). That is a way of saying that there should be no fundamental restrictions on recombination; i.e., that all laws of coexistence and restrictions on higher-level configurations should emerge from a low level ontology that recognizes no such constraints. And that in its turn is just an expression of a preference for organic explanation.

The preference for simple theories explains why we look for deeper, unifying models. Because if we can embed a motley collection of models, each recognizing degrees of freedom needed to support some local pocket of regularity, in a single model that reproduces all from a smaller number of degrees of freedom, we have simplified the number of *overall* number we need to recognize. Detectives at a crime scene use the evidence to form a theory about what happened in the case they are investigating. Why should they bother coming together pooling information and trying to form a single overarching theory? Because patterns and similarities across crime scenes present opportunities for organic explanation that will push them to a joint theory with a smaller number of overall degrees of freedom. If they can identify the perpetrator of one crime with the perpetrator of another, they will have taken steps at eliminating excess structure. This is the process that in science pushes us to search for a unified explanation of ever wider bodies of phenomena, an explanation that spans all of the special sciences taking the local pockets of established regularity into its sights, instead of conjoining them as a collection of disjointed theories. And it provides important clues about where theories with unnecessary complexity go wrong. They fail to recognize redundancy for what it is.

The preference for organic explanations can also provide insight into the difference between good and bad buck-passing explanations. Buck-passing explanations are unavoidable in science. We explain one event as the product of another, which in its turn gets explained by another, but at the end of the day, we are always left with events that are themselves unexplained. Some buck-passing explanations are clearly unsatisfying. The old children's song: "I Know an Old Lady" provides an extended and humorous example of buck-passing explanation in which a long succession of events is each explained by the preceding event in the chain, bottoming out in an unexplained event with little felt gain in explanatory value.[liv] On the other hand, some buck-passing explanations are clearly explanatory. When a doctor explains a wide range of a patient's symptoms by an underlying disease, one feels that something has been explained even if the disease itself has no known cause. I propose that the sense of explanatoriness in the second case is at least in part a product of the reduction effected in degrees of freedom. If a derivation from earlier conditions is not accompanied by a reduction in degrees of freedom, then the explanation is what I call *purely* buck-passing. If there is a reduction in degrees of freedom, then even though the explanation as a whole is buck-passing, some of the *regularity* in the explanandum has gotten explained organically along the way. Whenever we explain a disparate and wide-ranging collection of effects by deriving them from conditions that are manifestly less complex than the phenomena to be explained, it is the organic explanation of the regularities among the effects that leaves us with a sense of having achieved something.

We might also put this by saying that a buck-passing explanation is most explanatory to the degree that it is *unifying*, in the sense that a single degree of freedom is a common component in the explanation of a range of

effects.[lv] Copernicus defended his model of the universe in these terms. He writes (*De Revolutionibus*, Book 1, Chapter 10): "We thus follow Nature, who producing nothing in vain or superfluous often prefers to endow *one cause with many effects*." He shows how to explain the apparent motions of all of the planets using the Earth's motion around the sun as a common component, where Ptolemy has to replicate the Earth's motion relative to the sun separately for each planet. The result is that where there are many epicycles in the Ptolemaic system that replicate the relative motion of the earth and the Sun (a regularity that Ptolemy has to regard as coincidence) in Copernicus' system, these turn out to be the same motion, viewed from different frames of reference. Copernicus boasted repeatedly that his model was more *harmonious* than Ptolemy's and cited a number of further, specific instances of organic explanation as evidence.[lvi]

We can also connect unity in the above sense with David Deutsch's hard-to-vary explanations. Models that reproduce a wide and varied body of phenomena with a small number of degrees of freedom have the kind of hard-to-vary character that Deutsch points to as the characteristic of good explanations (and for him, the key to the universe).[lvii] They tend to be explanations in which everything is deeply interconnected so that any shift in one part of the system will affect all parts of the system.

Elegance is a more aesthetic notion, but one closely tied to this kind of unity. In a highly unified model, there is nothing arbitrary, inessential, nothing that can be added, taken away, or changed without affecting its connection to the phenomena. Einstein also connects unity to simplicity, and uses both in a sense that is measured by the number of degrees of freedom. He writes

"Although it is true that it is the goal of science to discover rules which permit the association and foretelling of facts, this is not its only aim. It also seeks to reduce the connections discovered to the smallest possible number of mutually independent conceptual elements. It is in this striving after the rational unification of the manifold that it encounters its greatest successes...."[lviii]

Einstein spoke often of simplicity as a kind of 'inner perfection' and although he professed not to know how to define it in more precise terms, he insisted that there was widespread agreement among scientists in judgments of comparative simplicity.[lix] In his own theorizing, he strove to weed out degrees of freedom not needed to reproduce discernible trajectories, pioneering the most powerful methods of excising idle wheels in the formalism.[lx] That he regarded these as explanatory successes and points to them as evidence of the great simplicity of his theory, connects simplicity to organic explanation. What Einstein refers to as the inner perfection of theory and what Copernicus meant by harmony is correlate with its capacity to provide organic explanations.

There is a lot more that could be said about each one of these points, and a good many other maxims of good explanation. But the overall upshot is that the preference for models with fewer degrees of freedom can serve as a unifying focus for a wide range of explanatory principles and theoretical practices.

Simplicity and truth

The question that remains, of course, is what *warrant* is there for using simplicity in this sense as an extra-empirical criterion for the construction of models? Note here, it is not just serving as a basis for choice among theories that have the same empirical content. Simplicity is guiding the choice of models that differ, often quite radically, in prediction. Copernican astronomy, the development of space-time theories from Newton through General Relativity, and the history of particle physics can all be motivated by a preference for increasingly simple models.[lxi] In each of these cases, audacious theoretical developments guided by largely theoretical preference for simplicity were vindicated by empirical test. The power of physical theorizing is that it allows us to go from little bits of information gathered under what are – from a cosmic perspective – quite special circumstances, and form inductive hypotheses that reach far and wide. Those inductive hypotheses can be tested to some extent and serve as the basis for further developments, and this bootstrapping cycle of observation-led speculation and speculation-led observation turns out to be a very powerful one. And modeling is an integral part of that process.

Einstein seems to have thought that the best one can do to justify the role of simplicity is a kind of meta-induction. He remarks that the physicist's conviction that "the totality of all sensory experience can be 'comprehended' on the basis of a conceptual system built on premises of great simplicity" will be derided by skeptics as a "miracle creed," but, adds, "it is a miracle creed which has been borne out to an amazing extent by the development of science".[lxii] He's certainly right that when science has looked for this kind of reconstruction, it has uncovered an order behind the phenomena that can be harnessed to increase precision and accuracy of predictions and guide intervention. It has been able to unlock the key to wider bodies of phenomena and explain what we do observe as the product of a relatively small number of ultimately differentiable components acting in accord with simple rules. But we would like some deeper understanding of why the method works. Is there some reason that we should expect it to work, not just to generate models that reproduce observed behavior in a compact and convenient form, but also to generate new predictions and guide interventions?

Schlick's answer was that "the greater simplicity of a theory depends on its containing fewer *arbitrary* elements" and that we should prefer the simplest theory because "we are then sure of diverging from reality at least no further than is necessitated by the bounds of our knowledge as such."[lxiii]. By 'arbitrary elements' he meant the excess elements in the formalism that are present in models recognizes more degrees of freedom in the system than are needed to reproduce the phenomena. Einstein reacted enthusiastically to Schlick's defense, and it goes well with intuition that I suggested at several points motivates preference for a simpler model which is that when a simpler model is available, the more complex model will be attributing more freedom of motion than we have evidence of. The simpler models do a better job of capturing compositional structure of the world up to level that can be discerned by macroscopic effects. They do a better of clearly, non-redundantly, and without overreaching the evidence reflecting the degrees of freedom that are implicated in producing macroscopic variation. This sort of justification shifts the burden of proof by basing the preference for simplicity not on a substantive ontological assumption about the simplicity of nature, but on epistemological conservatism. The justification is interesting because it is quite different from

justifications that are often foisted on the preference for simple models which involve substantive commitments to simplicity in nature, and suggests a different line of exploration than those emerging in the large body of literature on the connection between simplicity and truth.[lxiv]

Conclusion

The focus here has been on bringing into relief a type of simplicity that plays an important role in model-building. I started with some simple kinematical examples and tried to articulate the intuitive justification that favours simpler models. This is not the only type of simplicity that plays a role, but it unifies a surprising array of theoretical practices and explanatory principles, and focusing on it can be an important step towards a deeper understanding.

Adendum: **Objections and queries**

(i) Objection: the number of degrees of freedom is an abstract measure. It doesn't distinguish quite different accounts of compositional structure. Response; of course. It does not, and was not intended to, provide the sole criterion for the construction of a model. Questions about how degrees of freedom are divided into components and distributed among the parts of a model are difficult and have to be guided carefully by the details of the dynamics.

(ii) Objection: if this is to make any sense at all, there has to be a way of counting the number of degrees of freedom a theory recognizes (or, equivalently, the number of dimensions in its associated phase space) that is invariant under reformulation. But if we have a set of points representing the states of a system arranged in a 6 dimensional space, couldn't we just rearrange them along a line and now have a theory with a one-dimensional phase space, i.e., a theory that recognizes just a single degree of freedom in the system's state? Response: The answer is no, at least not in a way that preserves distances between points and the distances between points reflect features of the dynamics.[lxv] The number of degrees of freedom associated with a system by a theory is characterized by the dimensionality of the associated phase space, and that is an invariant quantity given the metrical constraints.

(iii) Objection; what of the phase spaces of infinite dimensionality? Response; generalization to the infinite case is not trivial. There is no general answer, but in some cases, there is a natural way of counting. In field theories, for example, every region of space contains an infinite number of degrees of freedom in the form of space-time points. But we can still compare theories by the quantities defined on space-time and we can compare the structure of space itself by the geometric objects defined on the manifold. The real trick to doing this right is to forget about counting the absolute number of degrees of freedom and look rather to starting with a phase with lots of structure and arriving at a reduced space by elimination. In that way we

can form an idea of what it means to lower the number of degrees of freedom associated with a theory without necessarily having a way of counting them absolutely.

(iv) Objection: if staying as close as possible to the evidence is the goal, why not stay with a simple description of the phenomena, neutral about underlying structure? If it is conservativeness we are after, that is the most conservative option. Response: staying as close as possible to the evidence is an auxiliary goal. The primary goal is the formation of powerful and accurate inductive hypothesis. This is the question of why we model at all. Why do we care about the compositional structure of the system? There are philosophers who deny that science is or should be in the business of produce a direct description of the categorical structure of the world. They think that it should restrict itself to providing a catalogue of regularities in the observable structure of the world, or perhaps its dispositions to produce certain kinds of manifest behavior. Mach was an early proponent of such a view, and Einstein had some early sympathy with. But he rejected it later on with his memorable dismissal of Mach's conception of a dynamics that restricted itself to the description of constant functional relations among sensations as a 'catalogue', not a theory. We have a practical interest in theories that are audacious in prediction and testability. Modeling is an essential part of the artful type of induction that doesn't just project observed regularities into the future, but leads to new kinds of phenomena, guides discovery and design. Knowing how to generalize patterns, the limits within which they hold, where they break down, what they depend on, being able not only to predict how a simple will behave, but intervene effectively in its behavior … all of that comes from modeling. The reason that we opt for the simpler model is that simpler models do better empirically. Any attributed complexity that goes beyond what is needed to reproduce the observed motions is either empty or runs the risk of messing up the inductive component of the theory, predicting phenomena that never occur.

(iv) Objection: you talk of parts and wholes and unrestricted composition. This doesn't fit quantum mechanics. What you've given is a very classical picture of ontology. Response; that is what is *wrong* with quantum mechanics. What we have in quantum mechanics is a theory that acts beautifully as a computational and predictive tool, but it does not furnish us with a clear and unambiguous ontology, not least because there are basic non-dynamical constraints on the joint states of complex systems that are built right into the rules by which we derive the phase space for complex systems from components. What is wrong with quantum mechanics from this point of view is that the particles that compose of a quantum system (in standard, non-relativistic quantum theory) do not behave like independent degrees of freedom. There is a lot to say here as well.

Can the scientific enterprise do without ontology? Doesn't the example of quantum mechanics precisely show us that although science has traditionally engaged in speculation about the ontological substructure of the world, it needn't? Perhaps. This is a fraught question. But to the extent that science is concerned with genuinely capturing the underlying categorical structure of the world, simplicity has a role.

[i] Zellner et al. 2001, p.2
[ii] We can get more precise about 'structure', but there is no reason to limit the generality. Decomposition into parts of any kind (subsystems, structural components, mechanisms, or material constituents) will count as a kind of structure. Ultimately, we may not have any way of characterizing structure except as whatever is usefully captured in a model. This kind of circularity is fine, for my purposes. There is no reason to restrict modeling at the outset of inquiry.
[iii] A complex dynamical system is any system that has parts (hence, 'complex') and evolves over time (hence 'dynamical').

[iv]See in motion http://www.officeplayground.com/Jupiter-Perpetual-Motion-12-inch-P1346.aspx

[v] I've separated the empirical component – data gathering – from the analytic component – discovering laws, but in practice they go together. We find regularities in the data that suggest laws, but data-gathering is guided by hypotheses about laws.
[vi] Since the balls maintain their integrity under evolution, it suffices to have a single variable that traces the position and momentum of the center of mass of each ball. If the balls changed color or shape, or if they divided like cells, we would need additional variables.
[vii] Of course, this isn't really a tag-team event; they work closely and simultaneously; in some sciences there is little division of labor at all. In physics, there is a fairly pronounced division.

[viii] last few pages of Einstein's "The Meaning of Relativity" (5th edition), in which he writes of his unified field theory: "*In my opinion the theory here is the logically simplest relativistic field theory that is at all possible. But this does not mean that Nature might not obey a more complex theory. More complex theories have frequently been proposed. . . In my view, such more complicated systems and their combinations should be considered only if there exist physical-empirical reasons to do so.*"

[ix] a non-redundant, complete phase space has exactly one point for each occupiable state.
[x]Even though the parts move in a three-dimensional space, their motion is restricted so that a single parameter representing the location of each along its orbit will suffice to specify its state.
[xi] Dynamical laws usually take the form of restrictions on trajectories through phase space. Though it should be noted that to say that histories are represented by trajectories already incorporates some restrictions on potential trajectories. A trajectory is a continuous path through phase space so only points that are contiguous in phase space will appear as succeeding states along any trajectory.
[xii] If we included momentum or velocity, we would have additional restrictions, e.g., points in which ball 1 is moving upward and ball 2 is stationary would be unoccupied.
[xiii] We very often find ourselves pushed to distinguish trajectories that cannot be observationally discriminated, but it is a recognized a point against a model when it does so.

[xiv] http://www.youtube.com/watch?v=4wLkF88toUc

[xv] http://www.youtube.com/watch?v=reg8NTAdFHE&feature=related
[xvi] Three times as many, if pixels varied in color, saturation, and hue.
[xvii] What is nice about the virtual examples is that they capture the idea that the underlying degrees of freedom need not themselves be localized in space.
[xviii] Any detectable or measureable effect a system has on another system under any conditions, so long as that effect is registered stably enough to be properly thought of in dispositional terms, counts as an observable. Dispositions can be probabilistic or categorical.

[xix] The distinction between Observables and Beables has connections to those other time-honored (and much disputed) philosophical distinctions; the distinction between data and theory, or appearance and reality. Although there's dispute over a general explicit definition, the distinction is easy enough to draw in particular cases and essential to any account of theorizing. For some of the dispute, see van Fraassen, B. C. (2001). "Constructive Empiricism Now", *Philosophical Studies* 106, 151–170, (2002). *The Empirical Stance.* New Haven: Yale University Press, van Fraassen, B. C. (2004). "Replies to the Discussion on *The Empirical Stance", Philosophical Studies* 121, 171–192, Churchland, P.M. (1985). The Ontological Status of Observables: In Praise of Superempirical Virtues, in P.M. Churchland and C. Hooker (eds.), *Images of Science. Essays on Realism and Empiricism with a Reply from Bas C. van Fraassen.* Chicago: University of Chicago Press, pp. 35–47.

[xx] Suppes 1962,"Models of Data", in Ernest Nagel, Patrick Suppes and Alfred Tarski (eds.), *Logic, Methodology and Philosophy of Science: Proceedings of the 1960 International Congress.* Stanford: Stanford University Press, 252-261. Reprinted in Patrick Suppes: *Studies in the Methodology and Foundations of Science. Selected Papers from 1951 to 1969.* Dordrecht: Reidel 1969, 24-35.
[xxi] In practice the generation of a data model and the activity of modeling aren't always clearly distinguishable, but it is useful to distinguish them.
[xxii] This is a bit of a cartoon, but a useful one to the extent that we can see different elements of the widely distributed and highly variegated activity involved in theorizing as contributing to different elements of this process. There is data gathering and the processing of data, the development of models meant to capture the contours of the data, and the testing of those models.
[xxiii] *On the Method of Theoretical Physics*, Philosophy of Science, Vol. 1, No. 2 (Apr., 1934), pp. 163-169. Einstein spoke often of simplicity, gave it a central role in his philosophy of science, and clearly intended his notion to be applied to the ontological component of a theory rather than its laws. A lot has been written about his views on simplicity. He himself struggled to find a fully

adequate explicit characterization of what simplicity was, though he gave many examples in his own work, and of all of the physicists who have written about simplicity, his notion comes closest to my own. Many of his remarks can be captured very well by reading simplicity as a measure of number of degrees of freedom, and the 'inner perfection' he thought was characteristic of a simple theory as the tight fit between its phase space and the allowed trajectories, but his views were complex. In some places he speaks of 'reducing the connections discovered to the smallest possible number of mutually independent conceptual elements'. "Although it is true that it is the goal of science to discover rules, which permit the association and foretelling of facts, this is not its only aim. It also seeks *to reduce* the connections discovered to the smallest possible number of mutually independent conceptual elements." In other places he spoke of it in logical terms, looking to the simplest set of axioms (or sometimes 'laws' or 'rules') from which the totality of natural phenomena could be derived." One can get no better discussion of Einstein than *The Shaky Game* (Fine, 1996), and Don Howard's work and Einstein's own writing. See also *Einstein: The Formative Years, 1879 – 1900,* (Howard and Stachel eds., 2000)

[xxiv] Per Bak, *How Nature Works: The Science of Self-Organized Criticality* (1996; pp.1-5)

[xxv] Bak, ibid..
[xxvi]Classically, the only degrees of freedom in how those components can be arranged were spatial, but there have been pro
[xxvii] A fundamental theory wants to isolate the ultimately separable components of matter. Any restrictions on configurations have to be explained. In a truly fundamental theory, the basic parameters should be genuinely distinct existences. If they really are separable components, there must be interactions in which they are separated. We can talk about electrons and protons because we can isolate those components and discriminate them from one another by their effects on measuring instruments, cloud chambers, and the like. If a theory is recognizing degrees of freedom that are *ultimately* inseparable (if, for example, it tells us x-ons are different from y-ons but there is no interaction in which x-ons and y-ons have different effects, then the theory is making distinctions that there is no empirical basis for.
[xxviii] Dirac 1930, p. 71, note 5.
[xxix] Ford K. 1963. "Magnetic monopoles", *Sci. Am.* 209: 122–31, p. 122
[xxx] There are actually two forms; either the Lagrange equations of the first kind treat constraints explicitly as extra equations. I am talking here about Lagrange equations of the second kind.
[xxxi] They are called 'generalized coordinates' as a holdover from when Cartesian coordinates were standard.
[xxxii] These sorts of preferences are defeasible. There are other considerations that may weigh against them.
[xxxiii] Incomputability of AIC and other information-theoretic measures of simplicity.
[xxxiv] See Forster, M. R. and Sober, E. [1994]:'How to tell when simpler, more unified or less ad hoc theories will provide more accurate predictions', *British Journal for the Philosophy of Science*, 45, pp. 1–35.

[xxxv]" I must confess herewith that I cannot at this point, and perhaps not at all, replace these hints by more precise definitions. I believe, however, that a sharper formulation would be possible. In any case it turns out that among the "oracles" there usually is agreement in judging the "inner perfection" of the theories." Einstein 1946, "Autobiographical Notes." In Schilpp 1949, 1–94. Quotations are taken from the corrected English translation in: *Autobiographical Notes: A Centennial Edition.* Paul Arthur Schilpp, trans. and ed. La Salle, Illinois: Open Court, 1979. pp. 21, 23.

[xxxvi]And we might take his own advice here to focus on his practice: "You wish to learn from the theoretical physicist anything about the methods which he uses. I would give you the following piece of advice: Don't listen to his words, examine his achievements." "On the Method of Theoretical Physics", Philosophy of Science, Vol. 1, No. 2, (Apr., 1934), pp. 163-169
[xxxvii]See *Essays on Symmetry* (Ismael, 2000), *Foundations of Space-time Theories* (Friedman, 1986), *Space, Time, Matter* (Weyl, 1952)
[xxxviii] Symmetry and Group Theory in Particle Physics (Costa and Fogli, 2011), Symmetries and Conservation Laws in Particle Physics (Haywood, 2010), Relativity, Groups, Particles (Sexl, Urbantke, and Urbantke 2000)

[xxxix]See Belot, G., "Principle of Sufficient Reason". "The standard phase space has a $3N$ dimensional configuration space and a $6N$ dimensional phase space, and is invariant under the six dimensional group of Euclidean symmetries. …We can …construct a *reduced phase space*, by taking the quotient of the standard phase space by the action of the Euclidean group. The points of the this space are equivalence classes of points of the standard phase space related by isometries, carrying a geometric structure inherited from that of the standard phase space … The resulting dynamical theory captures all of the invariant content of the standard theory." Another technique that deserves more than honorable mention, but that would take a more complicated discussion to introduce: Cosma Shalizi's computational dynamics.

[xl] Motion is just a stand-in for variation. Dimensions along which the state of a system can vary is just dimensions along which it can move in phase space.

[xli] The Humean Ban has played a variety of roles in philosophy, nicely documented in Wilson, J., "What is Hume's Dictum and Why Believe It?", *Philosophy and Phenomenological Research* 80 (3):595-637. Wilson also discusses a number of justifications for the ban. The justification for its role in physics is that the theorist looks to derive all regularities of coexistence ultimately from what she calls constitutional necessities. See Wilson, Jessica, 2010. "From Constitutional Necessities to Causal Necessities". In Helen Beebee and Nigel Sabbarton-Leary, editors, *Nature and its Classification.* London: Routledge; Stoljar, Daniel, 2007. "Distinctions in Distinction". In Jesper Kallestrup and Jakob Hohwy, editors, *Being Reduced: New Essays on Causation and Explanation in the Special Sciences.* Oxford University Press, Witmer, Gene, William Butchard, and Kelly Trogdon, 2005. "Intrinsicality without Naturalness". *Philosophy and Phenomenological Research*, 70:326–50, Armstrong, David, 1989. *A Combinatorial Theory of Possibility.* Cambridge: Cambridge University Press.

[xlii] A much stronger position would hold that our descriptive vocabulary has to be strong enough to individuate any entities we take ourselves to refer to, and so we literally can't speak of Beables that can't be discriminated by their effects. We don't need anything so strong here, and it would require a rather lengthy defense, but such a position would have the result that any theory that was more complex than it needed to be to reproduce descriminable movement had some unacknowledged redundancy in its phase space. I am attracted to the view, but the argument would be too complex and conjectural to defend.
[xliii] Regularities could, of course, be the product of coincidence or pre-established harmony. Pre-established harmony is either a divine doctrine, or derives a coincidence later on from a coincidence in initial conditions. To say that something is a coincidence can provide a kind of explanation (it can provide an answer to a 'why'-question), but I'll think of coincidences here *as unexplained* regularities. One doesn't need any complicated confirmation theory to understand the theoretical prescription to avoid coincidence. Coincidences are improbable conjunctions of events, and any improbability that attaches to coincidences will be inherited by the theory that needs them to explain phenomena.
[xliv] I toyed with saying they are explained by composition, but the notion of organic unity has connotations that are apt here; the idea that the explanation flows not from any external constraints imposed , but directly from the structure of the system itself.
[xlv] See Wilson, *Ibid*., "Intrinsically typed entities are (conditionally or unconditionally) necessarily connected by constitutional necessity just in case (i) the entities are not wholly distinct; and (ii) at least one entity constitutes the other. "

[xlvi] There are connections to other kinds of explanation, but the distinction here is more basic: unificationist approaches. kinematical explanation, mechanistic explanations and common cause explanations can be seen as forms of organic explanation.

[xlvii] It is tempting to say that the simpler a theory, the more work is done by organic explanation, but it would be hard to give an invariant measure of how much explanatory work is done by one means or another. These kinds of judgments are always partly dependent on what we consider significant. There is no way to put a theory-independent valuation over which of a pair of theories both of which have some explanatory success and some explanatory failures has the more significant successes. That is not a matter of counting, but a matter of what we regard as significant, and that is a practical matter. Scientists will have to use their own instincts about which research programs to pursue, but in pairwise comparisons of fully formulated theories that agree well enough with the evidence at hand, the simpler theory (and the one that recognizes fewer underlying degrees of freedom, the one that places the stronger organic constraints on observable trajectories) will be preferred as going less far beyond the evidence. The burden will be on the more complex theory to search for phenomena that demand the additional complexity.
[xlviii] The failure of parity in weak nuclear interactions is a nice example. When Pauli learned from Temmer of the experiments that showed the violation of parity, he refused to believe it. "That's total nonsense" , he reportedly told Temmer, and insisted that the experiment be repeated.

[xlix] See Grush, R., "Self, world and space: The meaning and mechanisms of ego- and allocentric spatial representation", *Brain and Mind*, for especially nice discussion of the complexity of those regularities and how they are incorporated into an allocentric model of space from which they emerge organically.

[l] Interventionism is one leading school of counterfactual account that has gained most acceptance in the scientific literature.
[li] E.g., the causal effects of smoking (as opposed to not-smoking) are those downstream effects that vary with smoking, when it is treated as a free variable (holding fixed auxiliary conditions that may remain tacit in informal contexts, but get made explicit in a causal model).

[lii]See especially: Reichenbach, H. (1956), *The Direction of Time*. Berkeley: University of California Press. Russell, B. (1948), Human Knowledge: Its Scope and Limits. New York: Simon and Shuster.
Salmon, W. (1971), "Statistical Explanation", in W. Salmon, *Statistical Explanation and Statistical Relevance*. Pittsburgh: University of Pittsburgh Press, pp. 29-88. Salmon, W. (1975), "Theoretical Explanation", in S. Korner (ed.), Explanation. Oxford: Basil Blackwell, pp. 118-45. Salmon, W. (1979), "Why Ask 'Why'? An Inquiry Concerning Scientific Explanation", in W. Salmon, *Hans Reichenbach: Logical Empiricist*. Dordrecht: Reidel, pp. 403- 25.
[liii] See Faye, J., Natures Principles, and Durham, I., Unification and Emergence in Physics: the Problem of Articulation", at http://fqxi.org/data/essay-contest-files/Durham_FQXi.pdf, and Mill, J.S., A System of Logic.

[liv] http://www.timmyabell.com/music/lyrics/ol/oldlady.htm
[lv] Note that this is not Phillip Kitcher's notion of unity, "Unification as a Regulative Ideal", *Perspectives on Science*, Volume 7, Number 3, Fall 1999, pp. 337-348. I register a promissory note to develop the contrast and push Kitcher towards my notion at some future time. Malcolm Forster's really nice work on this notion of unity, particularly his article on Mill and Whewell, http://philosophy.wisc.edu/forster/520/Whewell-Mill.pdf.

[lvi]E.g., that inferior planets are always seen in conjunction with the sun, and the retrograde motion of superior planets occurs when and only when the planet is in opposition to the sun. These are both emergent regularities that Ptolemy struggled to explain, but which flow from the geometry of Copernicus' system.

[lvii]http://www.ted.com/talks/david_deutsch_a_new_way_to_explain_explanation.html

[lviii] Albert Einstein, "Science, Philosophy, and Religion, A Symposium", published by the Conference on Science, Philosophy and Religion in Their Relation to the Democratic Way of Life, Inc., New York, 1941

[lix]" I must confess herewith that I cannot at this point, and perhaps not at all, replace these hints by more precise definitions. I

believe, however, that a sharper formulation would be possible. In any case it turns out that among the "oracles" there usually is agreement in judging the "inner perfection" of the theories." Einstein 1946, "Autobiographical Notes." In Schilpp 1949, 1–94. Quotations are taken from the corrected English translation in: *Autobiographical Notes: A Centennial Edition.* Paul Arthur Schilpp, trans. and ed. La Salle, Illinois: Open Court, 1979.pp. 21, 23. In some places he speaks of 'reducing the connections discovered to the smallest possible number of mutually independent conceptual elements', as in the quote above, in other places he spoke of it in logical terms, looking to the simplest set of axioms (or sometimes 'laws' or 'rules') from which the totality of natural phenomena could be derived.

[lx] "You wish to learn from the theoretical physicist anything about the methods which he uses. I would give you the following piece of advice: Don't listen to his words, examine his achievements." "On the Method of Theoretical Physics", *Philosophy of Science*, Vol. 1, No. 2, (Apr., 1934), pp. 163-169

[lxi] On Copernican astronomy, see Forster, *Ibid.*, On space-time theories, see especially, Friedman, M., *Foundations of Space-time Theories*, Princeton University Press, 1986: Princeton, NJ. On particle physics, see V. Mukherji and M Roy, American Journal of Physics -- December 1982 -- Volume 50, Issue 12, pp. 1100,"Particle physics since 1930: A history of evolving notions of nature's simplicity and uniformity"

[lxii] Einstein 1950, p. 342.

[lxiii] Schlick 1915, 154-155

[lxiv]There is a great deal to say about justification. One might suppose think that the intuitive justification has no value unless it can be formalized. The problem with trying to make such an argument is that we don't have a well-established theory of scientific evidence. What we do have is intuitions about which of a collection of theories compatible with a body of evidence receive the most support. Theories of confirmation try to systematize those intuitions and justify them with probabilistic arguments or otherwise, but the intuitions themselves are more solidly grounded than any theory of confirmation. although it is an interesting question whether the justification can be formalized in, for example, Bayesian or Likelihood terms, whatever support such formal theories of confirmation have rests on their ability to reproduce intuitive judgments, so justification runs in the other direction. My own interest in these questions was inspired by some closely related arguments in Clark Glymour's *Theory and Evidence.* For a Bayesian defense of the role of simplicity see, David L. Dowe, Steve Gardner and Graham Oppy, "Bayes not Bust! Why simplicity is no problem for Bayesians", at http://philsci archive.pitt.edu/2877/1/DoweGardnerOppy_draftBayesNotBust.pdf. for not specifically Bayesian discussions, see Kelly, K. (2004) "Justification and Truth-Finding Efficiency: How Ockham's Razor Works," *Minds and Machines*, 14, 485–505, in Forster, M. R. [1995]:'Bayes and bust: Simplicity as a problem for a probabilist's approach to confirmation.', British Journal for the Philosophy of Science, 46, pp. 399–424. Forster, M. R. [1999]:'Model selection in science: The problem of language variance.', *British Journal for the Philosophy of Science*, 50, pp. 83–102. Forster, M. R. [2000]:'Key concepts in model selection: Performance and generalizability', *Journal of Mathematical Psychology*, 44, pp. 205–231. Forster, M. R. [2001]:'The new science of simplicity', in A. Zellner, H. A. Keuzenkamp and M. McAleer (eds.) Simplicity, Inference and Model ling, University of Cambridge Press, pp. 83–119. Forster, M. R. [2002]:'Predictive accuracy as an achievable goal in science.', *Philosophy of Science*, 69, pp. S124–S134, URL http://philosophy. wisc.edu/forster/PSA2000.htm. Hannan, E. J. and Quinn, B. G. [1979]:'The determination of the order of an autoregression', *Journal of the Royal Statistical Society*, Series B (Methodological), 41(2), pp. 190–195. or less *ad hoc* theories will provide more accurate predictions', *British Journal for the Philosophy of Science*, 45, pp. 1–35. .

[lxv] The relevant theorem is due to Brouwer. Topological invariance of dimension: Suppose U and V are open subsets of n respectively m. If U and V are non-empty and homeomorphic, then $n=m$.