Why High-Level Explanations Exist

Michael Strevens

To appear in *Levels of Explanation*, OUP,
edited by Katie Robertson and Alastair Wilson

Abstract

High-level explanation would be impossible without a certain kind of independence of high-level behavior from low-level behavior. This partial autonomy has been characterized in various ways by scientists and philosophers; the present chapter advances a particular characterization deemed "semi-detachment". The main business of the chapter is to pose and suggest an answer to the question why semi-detachment should be so widespread, enabling the existence of explanatory scientific disciplines—evolutionary biology, sociology, economics—that operate almost entirely separately from the disciplines that study their lower-level substrates (microbiology and psychology, for example).

The approach explored here turns on low-level details' propensity to cancel or balance out in much the same way that the outcomes produced by a simple chance setup such as a coin toss balance out to produce stable frequencies. Such "canceling out" considerations, it is argued, provide at least a partial account of why high-level explanations in the sciences should be, not merely possible, but plentiful.

1. High-Level Explanation and Semi-Detachment

A simple evolutionary explanation invoking natural selection might go by way of a model structured as follows. The model contains variables representing the numbers of two variants in the relevant population, the fitness of each of these variants, and some assumptions about reproduction, including the way in which the variants are passed on from parent to offspring—in the simplest case, by an asexual mechanism in which the offspring has the same trait as its parent. Run the model, and the variant with the higher fitness comes to dominate the population.

More complex and interesting evolutionary models have more structure and more complications. Reproduction might be sexual, with the traits in question depending in perhaps elaborate ways on the genome. There might be many, or even an infinite number, of traits, if a phenotype can take on any of a range of characters (colors, shapes, etc.). The fitness of a trait might depend on the frequency of other traits in the population, or on other environmental variables subject to change in the course of the evolutionary process. Nevertheless, these more complex models have much in common with the simple model and with a vast array of other models used in high-level explanations.

On the one hand, they incorporate very precise representations of the systems whose dynamics they purport to capture. Evolutionary models may turn on tiny differences in fitness that slowly, but inexorably, work themselves out in favor of the fitter trait.

On the other hand, the precision in these representations falls well short of capturing the true causal intricacy of the system. Changes in a biological population are an aggregate of individual births and deaths, and these depend on such minute matters as the relative positions of predator and prey at particular times, which are in no way represented by the corresponding models. The models are in an important sense precise—they represent small differences, and what they predict and explain can turn on the details of those differences—but the precision is an entirely *high-level* precision. It manifests at the level of fitness, a parameter capturing a general fact about a phenotype in an environment, not a particular fact about the course of some specific organism's life as it makes its way around the environment. When such a model is put to explanatory use, the result is a paradigmatic high-level explanation.

Explanations having this high-level character proliferate in the sciences. An economic model of inflation may represent overall inflationary expectations, but not the expectations of individual economic actors, let alone their myriad other beliefs and goals. A model in population ecology, like many evolutionary models, may represent birth rates and death rates but not the facts that determine which organisms reproduce and when they die. A model of a chemically oscillating system, such as the Belousov–Zhabotinsky reaction, may represent the concentration of chemicals at each point in a shallow solution, yet leave out the motions and interactions of individual molecules which constitute the reaction. Even the most advanced weather forecasting models exclude a tremendous amount of causally relevant information. Many of these models manage passably accurate predictions all the same.

This chapter poses the question of how it is possible to have satisfactory explanatory models that operate mostly or wholly at the high level, omitting much of the causal detail that propels the systems in question along the trajectories that they take through the space of possibilities—omitting, that is, the eatings and matings of individual organisms, the decisions of individual buyers and sellers, the movements of individual molecules.

The answer to this question has two parts. The first part more carefully characterizes the independence or irrelevance of lower-level detail that makes high-level modeling feasible—a kind of autonomy that I call *semi-detachment* of the high level from the lower level. My characterization will not be especially formal or exacting, and indeed, it does not depart in any especially notable way from many other thinkers' attempts to describe a certain kind of autonomy a process may have from the minutiae of its causal implementation. (I am thinking here of the work of Garfinkel (1981), and more recently Batterman (2002, 2021), Woodward (2021), and Robertson (forthcoming), along with the remarks of scientists such as Simon (1996), the luminaries of the Santa Fe Institute (Cowan et al. 1994), and Goldenfeld and Kadanoff (1999).) Certainly, the differences between my characterization and that of, for example, Woodward or Robertson, will be of little importance in the execution of the second and rather more substantive part of this paper.

That second part sets out to provide an explanation of why semi-detachment is so widespread—an explanation that is all the more valuable because certain prima facie considerations suggest that it should be very rare indeed. Both for its intrinsic interest and because the practice of causal modeling in many disciplines hinges upon semi-detachment, this is a topic of immense significance. I would even go so far as to say that it is of comparable importance to, and has much greater practical significance than, philosophy's grand old problem of induction. Yet on the whole, it has received only fleeting attention from philosophers of science.[1] I hope to change that.

2. Semi-Detachment Characterized

A high-level behavior of a system is *semi-detached* from its lower-level foundations if the behavior can be predicted accurately using only properties of the system that are themselves high level. Additionally—since I am interested in prediction that is also explanatory—I require for semi-detachment that the predictive high-level properties are *differencemakers* for the behavior in question, and that the prediction proceeds by deriving the behavior as a causal consequence of these differencemakers. (I put aside, then, "models" that

1. I must make an exception for the case of universality in certain physical systems, memorably discussed by Batterman (referenced above) and his many commentators, and for the vast literature on the foundations of statistical mechanics (surveyed magisterially by Sklar (1993)). But the question of the prevalence of semi-detachment—not just in physics, but in the various sub-domains of biology, in psychology, and in the social sciences—has been raised and pursued by only a handful of philosophers, such as Strevens (2003, 2005), Loewer (2009), and Bhogal (this volume). Among general-audience writers, Cohen and Stewart (1994) and Gribbin (2005) have also emphasized the importance of the question.

make use of high-dimensional statistical correlations detected by machine learning to make their predictions.) The remainder of this section will offer remarks on this definition, along with glosses of important terms such as "level", "behavior", and "accuracy".

Let me begin with a couple of important observations about the notion of semi-detachment as a whole. First, semi-detachment is a property of a system's behavior, not of the system itself: one high-level behavior of a system might be semi-detached, another not. Second, semi-detachment is a kind of autonomy, makes possible a certain kind of multiple realizability, and has a kinship to notions of emergence such as that suggested by Wilson (2010). I name it using my own term of art to set it off from the many other senses of autonomy and emergence to be found in the philosophical literature. As observed in the previous section, it is nevertheless not something original to my own thinking; Garfinkel is an important precursor and Robertson and Woodward are fellow travelers. Third, there is inevitably a certain degree of vagueness and gradedness to the notion of semi-detachment. For simplicity's sake, however, I will tend to talk as though it is either present or not.

On to the notion of "levels". I don't assume any particular philosophical account, and certainly no particular metaphysics, standing behind talk of levels. We agree, I take it, that the population of a certain species in an ecosystem is a higher-level property of the system than the position of an individual member of the species. That sort of agreement is sufficient for what I want to say in this chapter.

Definitions aside, a few important remarks about levels. First, "high" and "low" are relative in my usage. The position of an individual organism is a low-level property in the context of modeling in population ecology, but not low at all in the context of statistical physics.

Second, as should be clear from the foregoing, when I talk about a high-level model, I mean a model in which not only the target behavior—the behavior that is supposed to be modeled—but also all other information about a system represented by the model concerns high-level properties. Such properties include, most significantly, statistical or aggregate properties of the system's parts: the number of organisms in a certain population or subpopulation in an ecosystem, the average kinetic energy per degree of freedom of the molecules in a chemical solution, the mean income of wage-earners in a certain age bracket and socioeconomic group, and so on. Other high-level properties are what you might think of as high-level background conditions: the ambient temperature, or hours of sunlight in the day, or central-bank interest rates.

Third, information about high-level properties is always at the same time about low-level properties, because the high-level properties of a system are in some sense—a sense that may vary with the property—determined by the low-level properties. It is the disposition of individual gas molecules that determines the pressure of a gas, for example, and the existence of individual members of a species that determines that species' population. To say that a model contains only information about high-level properties, then, is not to say that it contains *no* information about low-level properties, but rather that it contains only as much information about low-level properties as is entailed by the information it provides about high-level properties.

Fourth, throughout this chapter I assume that the high-level behavior of a system is ultimately determined by its low-level properties. Changes in the population of an ecosystem are brought about by, or depend on, the interaction of individual organisms: matings, eatings, and so on. The diffusion of one gas through another is brought about by the movements and collisions of individual gas molecules. And so on. For the purposes of this chapter, there is no need to understand this as a consequence of some entirely general reductionist principle (though I will admit that I am partial to such). It is quite enough that in the systems that I am most concerned with in what follows—especially biological populations and gases—it is true. In any case, the interest of the notion of semi-detachment, and thus the scope of the chapter, is limited to systems in which this low-level determination of high-level behavior is found. That is no great restriction, however, as such systems form the bulk of the subject matter of the special sciences.

What, next, is a "behavior"? Let me take my cue from science's model-builders: it is the sort of change (or lack of change) in a system that modelers build their models to predict or explain. To characterize behavior, then, we should look at modelers' predictive or explanatory goals. I don't intend to pursue that project very far, but I do want to emphasize an important consequence of this way of thinking about behaviors: modelers do not normally aspire to model high-level behavior (or perhaps any behavior) with exactitude. To put it another way, what they aspire to model is not exact behavior but approximate behavior.[2]

The inexactness can take several forms (here focusing for simplicity's sake on deterministic models). First, even when a model, on the face of things, traces an exact behavior, modelers do not take that behavior at face value. For the right gases in the right situations, it is appropriate to use the ideal gas model to predict or explain, say, the change in pressure that is caused by a certain change in temperature. The ideal gas model is exact: there is no limit to the precision it offers, if given precise input. But no one takes this precision seriously. Because the modeled gas is not ideal, we expect its behavior to deviate a little from the model's ostensible prediction, and such deviations are not considered predictive or explanatory failures. The model is treated as a success when its predictions are approximately correct.

Second, in many applications, modelers do not expect models to be even approximately correct on every occasion. The predictions of the ideal gas law, or for example Fick's laws of diffusion, may in principle deviate profoundly from a gas's actual behavior (though the probability of this happening in any particular instance is vanishingly low). The laws are nevertheless regarded as

2. For an overview of scientific modeling from a philosophical perspective, see for example Morgan and Morrison (1999),Weisberg (2013), and Frigg and Hartmann (2020).

excellent models of the systems in question.

In short, a predictively and explanatorily successful model of a system will typically capture that system's behavior only approximately, and may occasionally miss wildly.[3]

These remarks make it possible to characterize semi-detachment in somewhat more exacting terms. Like any attempt at philosophical precision, this one will not be free of artificiality and idealization, but I trust that it will bring a degree of clarity that justifies the cost. Let me suppose that predictive and explanatory models of high-level behavior consist of representations of possible states of the relevant system along with generalizations about how these states change over time. A simple model in population ecology, for example, might represent population levels, coefficients of reproduction and predation, a habitat's "carrying capacity" for each population, and a set of equations relating these properties so as to characterize the way that the populations will change over time. Let me suppose further that the generalizations are deterministic and exact, so that for any given specification of the states (the parameters and variables), the generalizations will issue an exact representation of the state of the system at any subsequent time.

Such a model, then, is a deductive system. The question of what it has to say about a system's behavior is a matter of interpreting the exact representations it makes of the system's state over time. For the reasons given above, a modeler will tend not to take such representations literally. If the model says that the population of rabbits in one year's time will be 300, the modeler might understand the prediction as follows: *very likely*, the rabbit population in one year's time will be *approximately* 300. The modeler's goals will dictate the extent of these tolerances, and therefore the circumstances under which they regard the model as satisfactory.

3. Because such models are not predictively perfect, it might be supposed that they are not explanatorily perfect, either. I do not rule out this possibility—that certain models that achieve greater accuracy by bringing in low-level detail are more explanatory than the high-level models that are my concern in this paper—but nor do I endorse it.

When the model performs well given the relevant tolerances, I say that it is predictively and explanatorily accurate. Accuracy, then, does not require exactly 300 rabbits; it requires approximately 300 rabbits, and the occasional complete miss is typically allowed. (Of course, we would not say that the model is accurate on that occasion; what I mean is that such mishaps are not inconsistent with saying that the model is accurate in general.) My proposed use of the term "accurate" is, I should perhaps add, merely an expository convenience.

To return to the definition presented at the beginning of this section, a high-level behavior of a system is semi-detached just in case there is a purely high-level description of the system that captures enough information about the behavior's differencemakers to model the behavior accurately.

The changes over time in the population of a certain ecosystem, for example, are semi-detached from the low level if it is possible to build a model, incorporating representations of only high-level properties such as population number, that accurately models those changes. Putting accurate initial conditions into such a model must, then, consistently result in a prediction that conforms at least approximately to the actual subsequent behavior of the system, in most instances.

Semi-detachment is quite independent of scientists' aims and beliefs. Though it is because of our modeling practices, and in particular because of our extensive reliance on black-boxing and our other uses of abstraction and idealization, that we have so great a need for semi-detachment, the existence of semi-detachment is not determined by our practices but rather by objective, worldly properties of the modeled systems themselves. Were these systems not to exhibit semi-detachment, a model that omitted certain low-level details would thereby omit details that make a critical difference to the behavior of the target system, and so would neither reliably predict nor fully explain that behavior.

Semi-detachment is therefore a great boon to science, indeed, an essential condition for the kind of high-level predictive and explanatory modeling without which the special sciences could not exist.

3. The Puzzle of Semi-Detachment

Semi-detachment, if not ubiquitous, is certainly not uncommon—so I have insinuated. That fact may seem as mysterious as it is convenient. Low-level details habitually have a high-level impact: the relative position of a certain fox and rabbit is exactly the sort of things to make a difference to the overall population of foxes and rabbits. Indeed, changes in population are determined by nothing but individual births and deaths, and therefore, it would seem, by the sort of fine-grained detail that high-level models of population ecology pass over in silence.

How, then, do things work out so happily? Why are our efforts at high-level modeling so often successful, both explanatorily and predictively? How does all the low-level detail, much of which causally contributes to high-level goings-on, conspire to cancel itself out, to add up, in all its causal potency and fecundity, to nothing? Or rather, nothing above and beyond its aggregate, its statistical manifestation at the higher level in terms of population numbers, average kinetic energy per degree of freedom, mean income, and so on?

A preliminary step in answering this question is to observe that the conception of a model's accuracy at the core of semi-detachment allows for the occasional gross deviation. If it is true that a flap of the butterfly's wings, or a twitch of the rabbit's ears, can cause a system to veer far from the trajectory that it would otherwise be expected to follow, then perhaps we can accommodate such paroxysms, if infrequent, under this escape clause. In effect, we are allowing that the high-level behavior we seek to capture with our models is an indeterministic behavior, a regularity marred by the occasional glitch. Such liberality makes it easier to find a suitable model.

But a certain tolerance on the modeler's part, though important, does not go to the heart of the matter. Exceptions aside, the modeler in their quest for accuracy demands high-level regularity, which is to say, behavior that can on the whole, if only approximately, be derived from high-level information alone. In order for there to be that kind of high-level regularity, the low-level aspects of a system's makeup that go unmentioned in a high-level model must either make no contribution whatsoever to a system's high-level behavior, or a contribution that is so consistent, so uncomplicated, that its net effect can be determined using high-level information alone.

The first option is not realistic, given that low-level details such as the relative position of individual predators and prey or gas molecules decide the difference between death and survival, collision and unimpeded travel, and these events in turn—deaths and collisions—are what drive changes in high-level properties such as an ecosystem's population or a gas's concentration. So it must be the second: all of that causally pertinent low-level complication and chaos somehow sums to a dependably rather simple ebb and flow at the high level.

The world need not have been so cooperative. Consider John Conway's Game of Life, a simple set of rules for cellular automata capable of generating a multifarious assortment of patterned and unpatterned behavior. Changing the state of a single cell in Conway's Game can send the system on a trajectory utterly different at even the highest level of description from the trajectory it would have traced without the change (figure 1).

Why is real life so much simpler than the Game of Life? Why does the pattern of population change in a habitat of rabbits and foxes depend only on a few high-level variables, rather than varying with the starting position of this rabbit or that fox? Why, in natural selection, does one trait that differs only slightly from another reliably outcompete the other, rather than the race's being decided by the spatial orientation of some particular organism or other when the trait first appears?

One simple and straightforward explanation is that we are cherry-picking. Even in the Game of Life, there are many regularities to be found. Gliders,

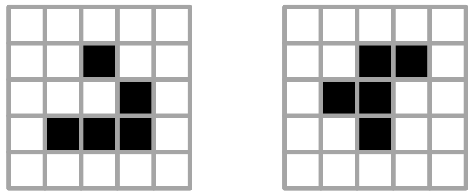

Figure 1: In Conway's Game of Life, the figure on the left, a "glider", keeps moving to the bottom right, cruising off to the edges of the universe (unless it encounters other life along the way). The figure on the right, the R-pentomino, if left to its own devices takes 1103 generations to settle down to a stable state. By that time it has sent off six gliders and is composed of 116 cells. Similar figures behave quite differently, and have quite different end states—or, like a glider, never settle down at all.

for example, move in a predictable direction at a predictable speed, provided that they do not encounter any other life. High-level models of the Game of Life may not be possible in general, but they are possible in certain carefully circumscribed circumstances.

Might our special sciences be selective and opportunistic in the same way? Might high-level modelers with a nose for semi-detachment converge on the few places where it is found, giving the impression that the property is widespread when in fact it is rare but well attended? Might the roving spotlights of scientific research pick out semi-detached high-level behaviors not because, wherever you point the light, you will find semi-detachment, but because the lights are manipulated so as to pick out nothing else?

Scientists will ever, of course, incline toward low-hanging fruit. A close look at high-level modeling practices suggests, however, that such bounty is remarkably common. Consider, for example, the way that students are taught to apply the models of statistical physics, not tentatively and selectively but enthusiastically and indiscriminately. Statistical models are simply expected to work in any of the enormous range of systems to which they ostensibly apply—and on the whole, they do.

Perhaps even more striking, because of the structural complexity and diversity of the systems involved, is the widespread applicability of high-level modeling strategies in biology. The models of population ecology are deployed wherever there are populations to model—whether to represent foxes preying on rabbits or malaria parasites on human children. The models of population genetics, similarly, are put into action wherever evolution is going on.

It cannot be said that these models invariably succeed as predictors the first time around. But when they fail, the modeler's assumption of semi-detachment is virtually never questioned. Instead, it is assumed that the model in question does not contain enough high-level information: ecological models may therefore be enhanced by building in representations of sub-populations, such as different age groups, or population genetics models by building in representations of more complex types of genetic interaction or mating preference. Modelers facing difficulties, in other words, double down on semi-detachment, proceeding as though models that contain sufficient high-level precision and detail will in the end attain some measure of predictive adequacy. This is, if anything, the reverse of cherry-picking: the modeler is not scanning the world for a few choice opportunities to apply their favored technique, but rather choosing to apply their technique anywhere and everywhere with an almost guileless confidence that is, in fact, rewarded over and over.

In short, while it is true that a great number of physical and biological systems have yet to be tested, those to which high-level modelers have applied their craft have tended, on the whole, to yield to the high-level approach—not because modelers are scrupulously selecting their targets, but because their habitual expectation of semi-detachment is satisfied far more often than not.

4. Explaining Semi-Detachment

4.1 Canceling Out and the Statistical Approach

Let me introduce the canceling-out explanation of semi-detachment by looking at simple monatomic gases. The molecules in a gas confined in a box careen around in the most chaotic way, sending each other flying in every direction. The overall state of the gas consists in nothing more than the positions and velocities of these molecules. Yet the aggregate of this intensely disordered motion—the high-level behavior of the gas—is a paradigm of order. A gas released into a box will almost immediately settle into an equilibrium state in which its molecules are approximately evenly distributed through the available space, and the speeds of the molecules conform approximately to the Maxwell-Boltzmann distribution (figure 2). Shift a molecule to the right as you release the gas and (almost certainly) it will make no difference: the gas will head to the same equilibrium state and stay there for as long as you care to watch.

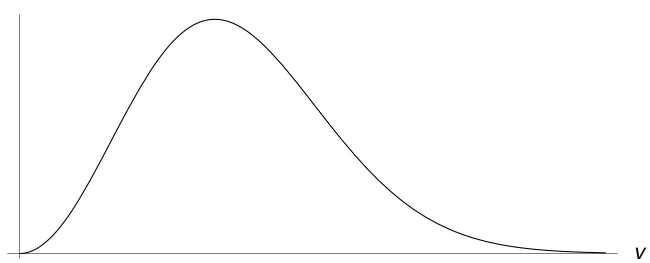

Figure 2: The Maxwell-Boltzmann distribution, showing the distribution of speed (i.e., velocity magnitude) for a gas at equilibrium

The reason that low-level chaos adds up to high-level simplicity is revealed by kinetic theory: the chaos (with very high probability) cancels out. A small change in the position of a molecule can, and usually will, result in that and many other molecules' later having completely different positions and velocities from those that they would have otherwise had. But from the high-level perspective, such changes are mere fluctuations, which in the aggregate are overwhelmingly likely to more or less balance each other in a way that results in a steady mean whose value is quite independent of the low-level details.

It is possible to understand the basic mathematical principle at work without plumbing the depths of statistical mechanics. Think of a coin tossed many times in succession. After a certain initial period, the frequency of heads will settle down to an equilibrium state in which it fluctuates, very slightly, around one-half. The outcomes of the tosses are about as disordered as they could be, but that very disorder means that they balance out so as to ensure (with just the slightest chance of sizable deviation) a stable high-level behavior.

In a gas, the changes in molecular position and velocity that arise from motion and collision cancel each other out in the same sort of way. For every molecule that makes its way toward the left-hand side of the box, there is very likely another that makes its way toward the right-hand side, and so on, so that the overall distribution of molecules throughout the box remains the same. Likewise, for every molecule that, thanks to a collision, undergoes a sudden increase in speed, there will very likely be one that undergoes a sudden decrease in speed so as to replace it in the overall speed distribution shown in figure 2. (The Maxwell-Boltzmann distribution is the equilibrium distribution precisely because it uniquely equalizes the rates of outflow from and inflow to each point along the speed axis.)

In short, a promising approach to explaining semi-detachment invokes the canceling out of otherwise causally relevant low-level details, in the way described by statistical models of coin-tossing and the kinetic theory of gases.

That observation prompts two questions. First, to what extent can we explain semi-detachment in other kinds of systems, such as ecosystems, using this approach? Second, even where technically feasible, are these explanations any good? They may work mathematically, but do they capture the real reasons for semi-detachment?

The answer to the first question is an unequivocal yes. I will argue by example. Consider a simple Lotka-Volterra model of a predator/prey ecosystem, representing the way in which interacting prey and predator populations—foxes and rabbits, say—change in time.[4] The model contains just two variables, representing the populations in question—x for the number of prey and y for the number of predators. It is encapsulated in two differential equations:

$$\frac{dx}{dt} = Rx - Pxy$$

$$\frac{dy}{dt} = Qxy - Sy$$

The first equation states how the rate of change of the population of prey ($\frac{dx}{dt}$) depends on the current population of prey and predators, as well as two constants R and P. The equation's first term Rx represents the rate of increase in the prey population due to reproduction; the constant R, then, is proportional to the mean number of offspring produced per member of the prey population. The second term represents the rate of decrease in the prey population due to predation. That number is proportional to the population of both predators and prey: the more prey milling about, the more are taken by any individual predator, and the more predators on the prowl, the more the total number of prey taken—or so the model asserts.

The second equation states how the rate of change of the population of predators depends on the current population of predators and prey. The first term is the rate of increase in the predator population due to reproduction, conceived of as proportional to the available food, and therefore to the rate of predation—the flip side of the second term in the first equation. The second term represents the rate of decrease in the predator population due to death from any causes. The constant S therefore captures the predator death rate per capita.

4. For a philosophically inflected discussion of this model and its uses, see Weisberg and Reisman (2008).

Such elementary models typically cannot predict with much accuracy changes in the modeled populations, but they can predict and explain qualitative features of population change—for example, the tendency of many ecosystems to manifest a robust equilibrium, in which prey and predator populations are roughly constant, or the tendency of a general biocide (a deleterious change in the environment that equally kills predators and prey) to result, in the medium term, in a larger relative proportion of prey. With respect to certain qualitative high-level behaviors, then, the Lotka-Volterra model is accurate in my sense. (More sophisticated high-level models, I should add, can do a much better job of quantitative prediction, and are used as forecasting tools by park administrators, fisheries managers, and so on.)

The accuracy of Lotka-Volterra models illustrates the semi-detachment of many high-level behaviors in a simple ecosystem. The qualitative effect of a general biocide, for example, depends only on the aspects of the modeled system that are explicitly represented in the model. It does not depend, then, on low-level details such as the positions of particular organisms (or, indeed, on many high-level aspects of the system).

How to understand this lack of dependence, this semi-detachment? The mathematics of the Lotka-Volterra model itself is deterministic through and through. Nevertheless, we can explain semi-detachment as a matter of low-level causes "canceling out" by conceiving of the underpinnings of the behavior described by the model statistically, that is, along the same lines as we conceive of coin-tossing or the statistical mechanical underpinnings of the physics of gases.

Consider, for example, the predator death term in the second equation, namely, Sy. It is rather natural to interpret the constant S as representing the probability that any particular predator dies over the course of a time interval of length 1 (in the units in which the model measures time). You might think of it this way: as a predator makes its way through the system, it has a certain chance of contracting a fatal infection or suffering a fatal accident.

It's a random matter, rather like drawing balls from an urn. Pick the black ball, and it's game over.

Drawings from an urn are like coin tosses. Over time, black balls will be drawn with a frequency that represents their proportion in the urn. To predict whether or not the next ball to be drawn will be black, we need a compendious description of the state of the urn and fearsome amounts of computation. But to predict the frequency with which black balls will be drawn, we need only know the proportion of balls that are black. The remainder of the facts about the state of urn impact the frequency only as fluctuations that, in the medium term, with high probability cancel out.

Likewise for the ecosystem: though to predict the death of any particular organism, we might need to know an immense amount about its position, condition, and the positions and conditions of every other organism in the system, to predict the overall rate of death we need only a relatively small amount of high-level information, determining the overall degree of danger. All of the hurly-burly of everyday life can, then, be compressed into a single number analogous to the urn's black ball proportion, a high-level property of the system as whole that accurately models the death rate. Everything else affects the death rate only as a series of fluctuations that very likely, if the population is not too small, balance out.

Or consider the term in the first equation that represents the rate at which prey are eaten, namely, Pxy. We can think of the vicissitudes of predation in the following way. Predators roam the habitat, looking for prey. The more prey there are, the more likely a predator is to encounter one and to eat it. It is as though each predator is drawing balls from an urn, determining whether or not they find a prey in some particular patch of the habitat. The urn contains a ball for each patch, either fleshy pink for "prey" or white for "empty". The more prey there are in the system, the higher the proportion of pink balls. Each predator, then, will have a chance of catching a prey proportional to the number of prey. Now, this sampling is going on for every predator. The expected number of prey caught in a given time interval, then, is proportional to the number of predators (the number of animals sampling the urn) and the number of prey (the number of pink balls in the urn, given that the total number of balls—representing approximate positions in the habitat—is fixed).

Over time, and if the populations are large enough, the actual rate at which prey are caught will very likely closely track the expected number. Thus it will be equal to a constant multiplied by the predator and prey populations—or Pxy, as in the model. The low-level complexity inherent in the biological dynamics of predation contributes to the predation rate only, again, in the form of fluctuations, much like the fluctuations in the frequency of "pink" caused by individual urn-drawings, canceling out and leaving behind a rate of death by predation that is dictated by high-level properties of the system alone.

The same kind of story might be told for any high-level model of this sort—any high-level model consisting of difference or differential equations representing change as a function of high-level quantities, when that change is driven by the outcomes of numerous low-level events such as predatory pounces or molecular collisions. We have a general strategy, then, for understanding the semi-detachment of high-level behavior—not, perhaps, applicable to every model in the special sciences, but apt for a vast range, including models of natural selection, economic equilibrium, statistical physics, and more.

4.2 Are Canceling-Out Explanations Valid?

The strategy is only as good, however, as its assumptions. Are the foragings and ultimate fates of predators like urn drawings or coin tosses in the relevant way? There are some strong prima facie reasons to think not, and thus to doubt the canceling-out explanation of semi-detachment, answering my second question above—is the explanation any good?—in the negative.

An initial concern is that the statistical strategy ascribes objective probabilities where there are none. Arguably, the low-level dynamics of predator-prey interactions are nearly deterministic, in the sense that any fluctuations bubbling up from the quantum level have little or no effect, characteristically, on their outcomes. And where there is determinism, some would say, there can be no physical probability or "chance" (Schaffer 2007).

But this concern should not delay us. The low-level dynamics of a coin toss are known to be nearly deterministic, yet a statistical model provides a satisfying explanation of coin-tossing's tendency to deliver a stable frequency in the medium to long term, showing how low-level detail cancels out. This may not be physical probability or "chance" in a grand metaphysical sense, but our goal is not metaphysical enlightenment but physical understanding, and statistical modeling is a fine way to understand canceling out and stable frequencies in deterministic systems, as in the coin toss and as in my interpretation of the underpinnings of the Lotka-Volterra model.[5]

Some more telling objections to the canceling-out explanation raise concerns not about the strategy of statistical modeling as such, but rather about the particular statistical assumptions that go into the story. I understood the predator death term Sy in the Lotka-Volterra model as a consequence of a fixed probability that any predator in the system would die over the course of the relevant interval of time (one unit of the time variable t). But surely it is not the case that every predator has an equal chance of dying in a given span of time. Very young, very old, or very sick organisms are more likely to go under.

This is all true. A better understanding of the predator death term is as follows. Assume that the age and health profiles of the population are in equilibrium. The proportion of organisms that are very young, then, remains the same even as the absolute number of organisms changes, as does the proportion of organisms that are old or sick. Then we can understand the

5. As argued by Strevens (2003), Myrvold (2021), and others, following a tradition initiated by von Kries (1886) and Poincaré (1896).

constant S in the death term not as a single death probability valid for every organism, but rather as a weighted mean of such probabilities—in technical terms, the marginal probability of predator death. Given the assumption of age and health equilibrium, the marginal probability will not change (unless the underlying probabilities change), and so we can safely put it to work to model the predator death rate.

What if, in a given ecosystem, age and health are not in equilibrium? Then the simple model examined above is not valid. What we need instead is a model that tracks the age and health structure of the population, with separate variables representing the number of young, adult, and old predators, sick and healthy predators, and so on for any factor that affects the probability of death, along with coefficients representing the differing probabilities. And population ecologists do indeed employ such models when necessary. Observe that like the original model, they represent the system in wholly high-level terms: the number of young predators is as much a high-level property of the population as its total number. The behavior of the system characterized in these somewhat finer-grained terms, then, is a high-level behavior, and one that floats free of further low-level detail—one that is semi-detached.

A second concern is that the probability of death for a particular organism depends on low-level details that, by contrast with age and health, cannot be packaged into high-level statistics in the way just proposed. Some predators are killed by falling trees. You might suppose, then, that an organism's chance of death will depend on its proximity to dead or dying trees. As it gets closer to such a tree, the probability of death inches up (ceteris paribus). Likewise, the probability of a prey's death surely depends on its proximity, at any moment, to a hungry predator. In short, death probabilities depend on the details of relative positions at particular times, information that is never represented, even statistically, in high-level population models.

This variation in the probability of death might be handled using something similar to the equilibrium posit above—a assumption, that is, that each prey spends about the same proportion of its time near and far from hungry predators, and that each predator spends about the same proportion of its time near and far from weak-rooted trees. That seems a promising strategy. It is, after all, on the whole a matter of pure chance whether an organism finds itself in this sort of peril: there is no reason why one organism should receive any greater exposure to danger than any other (putting aside phenotypical differences that could be captured by a population structure model).

A more rigorous (if far from deductive) argument to this effect is developed in Strevens (2003), chapter four. There I develop the idea that very small, short-term fluctuations in a creature's day-to-day meanderings will, thanks to various kinds of sensitivity to initial conditions, have a randomizing effect on matters such as proximity to danger. It really is as though organisms of a given sub-population in a given habitat are all making drawings from the same urn, subject to the same probability of drawing the red ball of peril.

Surprisingly, then, it turns out that the chaotic aspect of life—the sensitive dependence of important outcomes on small, seemingly insignificant matters of fact—which was portrayed as a threat to semi-detachment in section 3, is an important factor in ensuring the canceling out that makes semi-detachment possible. It is not the only factor, or else the Game of Life would yield to the high-level approach, but it is an essential part of the story.

A final concern with the statistical thinking behind the canceling-out approach is that it assumes stochastic (i.e., statistical) independence where there might seem to be none. Return for a moment to the case of coin tossing. Each toss in a series is stochastically independent of the outcomes of the other tosses, meaning that the other outcomes make no difference to the probability of heads on that particular toss: it is one-half whether preceded by a head, a tail, or some longer and more elaborate sequence.

Stochastic independence is a part of what explains the tendency for outcomes on a series of coin tosses to cancel out, resulting in a frequency for heads that in almost every case fluctuates just very slightly around one-half.

If the canceling-out explanation of semi-detachment is to be applied to the behavior represented by the Lotka-Volterra model, then, a corresponding independence assumption is required. The probability that one predator dies should be independent of the death of another; the probability that one prey is eaten should be independent of whether any other prey suffers the same ending.

Such an assumption may appear dubious, for two reasons. First, in the case of the tossed coin and other such randomizing devices, the rationale usually given for supposing stochastic independence is causal independence: the probability of heads on a given toss does not depend on the outcomes of other tosses because it is causally disconnected from the others. That certainly does not hold in the ecosystem: prey are continually interacting with each other and with predators, and vice versa.

Second, there are positive reasons to think that certain of these causal dependencies undermine stochastic independence. A predator can eat only so much at a time. If one prey is consumed, then, the probability that another will be consumed shortly afterward surely decreases, with one less hungry animal on the prowl.

In fact, however, an assumption of *approximate* stochastic independence—that is, outcomes' having very little if not zero impact on other probabilities—can be sustained in the ecosystem. It is true that in many cases, a predator's feeding will make all the other prey a little safer. But in an ecosystem of any size (and we need the size in any case to get canceling out), the effect is small. The predation terms in the Lotka-Volterra model remain roughly correct.[6]

That accounts for one possible source of dependence. But the interaction between the animals in an ecosystem is so pervasive, so intense, and so potentially consequential—I remind you again that very small variations in

6. To put it another way, it is as though the model supposes that the notional urn-drawing is with replacement (a ball is put back in the urn immediately after it is drawn), when in fact it is not. If an urn contains sufficiently many balls of each color, however, the difference between sampling with and without replacement is on average quite small.

position can make the difference between life and death—that it is impossible to maintain that the creatures persist in a state of even approximate causal independence. Their causal connections are thick and tangled; there is virtually no causal independence whatsoever. (Here sensitivity to initial conditions resumes its former role as a spoiler.)

Causal independence may be (at least in normal circumstances) sufficient for stochastic independence, but in other work I have shown that it is far from necessary. Consider two tossed coins that collide in midair. They interact substantially—certainly enough to make a difference to whether the coins land heads or tails. Yet under a wide range of conditions, the outcomes are stochastically independent. Learning how one coin lands gives you no help in predicting how the other lands (Strevens 2015). Further, a strong case can be made that the relevant probabilities in an ecosystem work in much the same way. (Strevens (2005) gives an accessible visual treatment; Strevens (2003) goes deeper.) The canceling-out explanation is therefore viable after all.

4.3 Expanding the Circle

I've said quite a bit about population change in ecosystems, but what about other behaviors and other kinds of system? Economic systems? Chemical systems? Social systems of various stripes?

The canceling-out approach is applicable—generalizing from the cases discussed above—if the following conditions hold. First, the semi-detached high-level behavior in question involves high-level properties whose dynamics are determined by the outcomes of many individual low-level events (births, deaths, changes in molecular velocity, coin tosses). Second, these low-level outcomes have probabilities whose values depend only on high-level properties (or fixed features) of the system in question, as the probability of prey death depends only on the overall number of predators and prey (and perhaps a fixed distribution of age and health). Third, the low-level outcomes are at least approximately stochastically independent.

That is a sufficient condition; perhaps something along the lines of a canceling-out explanation can be given in related cases as well. Or perhaps in some systems, the canceling-out approach will constitute one part of a multi-pronged strategy. For social systems, for example, canceling-out might explain certain regularities in the social background which, in tandem with the plasticity of human behavior, account for further high-level regularities in the course of human affairs.

In other cases, the best explanation of semi-detachment may have nothing to do with canceling out. If we want to understand semi-detachment across the board, then—if we want to understand the viability of high-level modeling and the power of high-level explanation wherever they are found in the special sciences—we will need a veritable toolkit of techniques. Canceling out, however, will surely constitute one of this toolkit's principal explanatory instruments.

References

Batterman, R. W. (2002). *The Devil in the Details: Asymptotic Reasoning in Explanation, Reduction, and Emergence*. Oxford University Press, Oxford.

———. (2021). *A Middle Way: A Non-Fundamental Approach to Many-Body Physics*. Oxford University Press, Oxford.

Cohen, J. and I. Stewart. (1994). *The Collapse of Chaos*. Viking, New York.

Cowan, G., D. Pines, and D. Meltzer (eds.). (1994). *Complexity: Metaphors, Models, and Reality.* Addison-Wesley, Reading, MA.

Frigg, R. and S. Hartmann. (2020). Models in science. In E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*. Spring 2020 edition. URL = <https://plato.stanford.edu/archives/spr2020/entries/models-science/>.

Garfinkel, A. (1981). *Forms of Explanation*. Yale University Press, New Haven, CT.

Goldenfeld, N. and L. P. Kadanoff. (1999). Simple lessons from complexity. *Science* 87:87–89.

Gribbin, J. (2005). *Deep Simplicity: Bringing Order to Chaos and Complexity*. Random House, New York.

von Kries, J. (1886). *Die Principien der Wahrscheinlichkeitsrechnung*. Mohr, Freiburg.

Loewer, B. (2009). Why is there anything except physics? *Synthese* 170:217–233.

Morgan, M. S. and M. Morrison (eds.). (1999). *Models as Mediators: Perspectives on Natural and Social Science*. Cambridge University Press, Cambridge.

Myrvold, W. (2021). *Beyond Chance and Credence: A Theory of Hybrid Probabilities*. Oxford University Press, Oxford.

Poincaré, H. (1896). *Calcul des Probabilités*. First edition. Gauthier-Villars, Paris.

Robertson, K. (Forthcoming). Autonomy generalised; or, why doesn't physics matter more? *Ergo*.

Schaffer, J. (2007). Deterministic chance? *British Journal for the Philosophy of Science* 58:113–140.

Simon, H. A. (1996). *The Sciences of the Artificial*. Third edition. MIT Press, Cambridge, MA.

Sklar, L. (1993). *Physics and Chance*. Cambridge University Press, Cambridge.

Strevens, M. (2003). *Bigger than Chaos: Understanding Complexity through Probability*. Harvard University Press, Cambridge, MA.

———. (2005). How are the sciences of complex systems possible? *Philosophy of Science* 72:531–556.

———. (2015). Stochastic independence and causal connection. *Erkenntnis* 80:605–627.

Weisberg, M. (2013). *Simulation and Similarity: Using Models to Understand the World*. Oxford University Press, Oxford.

Weisberg, M. and K. Reisman. (2008). The robust Volterra principle. *Philosophy of Science* 75:106–131.

Wilson, J. (2010). Non-reductive physicalism and degrees of freedom. *British Journal for the Philosophy of Science* 61:279–311.

Woodward, J. (2021). Explanatory autonomy: The role of proportionality, stability, and conditional irrelevance. *Synthese* 198:237–265.