

Statistical Autonomous Explanations and the Patterns of Nature: A Modified Account

Travis Holmes and André I. Ariew

Abstract

This paper aims to provide a fuller account of purely statistical pattern-level explanations –i.e. those which explain macro-level events by invoking limit theorems. The statistical autonomous explanation (SAE) account is modified via integration with the maximum entropy approach for generating limit distributions. This achieves two important results: (1) the range of SAEs is vastly extended and shown to range over many different kinds of limit distributions; (2) the modified account permits answers to questions about why these limit distribution patterns are so common in nature; why these patterns are robust; and also, why these patterns are insensitive to most lower-level details pertaining to the characters or events which comprise the statistical ensemble. The modified account can be understood as a corrective for many extant accounts of statistical pattern-level explanations which fail to answer these crucial questions.

Keywords: Explanation; Limit distribution; Normal distribution; Statistics; Entropy

Table of Contents

1. *Introduction*
2. *Statistical Autonomy and the SAE Account of Statistical Pattern-Level Explanations*
 - 2.1 *The notion of statistical autonomy*
 - 2.2 *Statistical autonomous explanations*
3. *Modifying the SAE Account: Limit Distributions and Maximum Entropy*
4. *The Modified SAE Account in Application: Case Studies*
 - 4.1 *Case study 1: Genetic drift*
 - 4.2 *Case study 2: Earthquakes*
5. *Conclusion*

1. Introduction

In 1877, Francis Galton delivered a lecture entitled “Typical Laws of Heredity” to the Royal Institution in London whose purpose was to explain a curious pattern of heredity detected at the level of whole populations and over the course of generations. As he described it:

In each generation there will be tall and short individuals, heavy and light, strong and weak, dark and pale; yet the proportions of the innumerable grades in which these several characteristics occur tend to be constant (Galton [1877], p. 1).

The constant was the well-known “law of error” (which he called the “law of deviation”), a probability function previously discovered in a range of phenomenon from outcomes of games of chance to observational data to heights and chest sizes of soldiers. Its graphical representation is a bell-shaped probability distribution with a peak at the mean and slopes that trace the variance or spread around the mean. The phenomenon Galton wished to explain is why do the processes of heredity—assuming external conditions are constant—conform pretty nearly to the curve of deviation. The hereditary phenomenon has consequences for the efficacy of Darwin’s natural selection. It demonstrated that a central assumption of Darwin’s theory was wrong, that individuals do not “equally tend to leave their like behind them” (*Ibid*,

pg. 2) Consequently, selection for mean characters will not eliminate the extremes in the population, for the processes of heredity tends to preserve them. So, for example, children and grandchildren of tall parents tend not to be as tall; instead, they tend to revert to the mean height of the population.¹

Galton illustrated the law of deviation “in action” with an apparatus that he created, a shot-dropping machine resembling a pachinko game, with rows of pins arrayed in a quincunx pattern (think of the face of the five on a six-sided die) and vertical compartments at the bottom that captured the piles of shot. Galton explained that “while the courses of the pellets are extremely irregular...you will observe the regularity of the outline of the heap formed by the accumulation of the pellets” (*Ibid*, p. 5). Runs of the quincunx explain how the distribution pattern emerges from the aggregate of a variety of micro-level events: “the collective actions of a host of independent petty influences in various combinations” (*Ibid*., pg. 5). Yet, as Galton was quick to point out, explaining how micro-events aggregate to form the curve of deviation is not the same as explaining how the whole distribution curve is maintained by the processes of heredity over the course of generations. Galton writes: “The conclusion is of the greatest importance to our problem. It is, that the processes of heredity must work harmoniously with the law of deviation, and be themselves in some sense conformable to it” (*Ibid*, pg. 8)

In his lecture Galton sketched some potential answers, but he didn’t arrive at his famous and surprising answer until 1889. The maintenance of the bell-shaped stable distribution of characters over the course of generations is a mathematical consequence of the normal law. If a population is normally distributed in one generation (and external conditions are constant) it can be deduced from the normal law that the population of the next generation will also be normally distributed around the same mean and with the same measure of dispersion. What makes Galton’s explanation so surprising is that it refers to a mathematical idealization as opposed to an extant biological process. No real-world population is normally distributed because no real-world population is infinite in size and random in its sampling (that is what makes the explanation an idealized one). Nevertheless, large populations approximate the normal distribution. Even in 1877 Galton recognized the novelty of the kind of explanation he presented: “The typical laws are those which most nearly express what takes place in nature generally; they may never be exactly correct in any one case, but at the same time they will always be approximately true and always serviceable for explanation” (*Ibid*, pg. 17).

Galton’s novel use of statistics to explain a real-world phenomenon led him to develop the statistical techniques of correlation, linear regression, and a variety of standards for goodness of fit between data and theory, all still used by scientists. According to historian Stephen Stigler, Galton’s work “represents the most important step in perhaps the single major breakthrough in statistics in the last half of the nineteenth century” (Stigler [1990], p. 281). Further, Galton’s mathematical theory of inheritance was the basis for the 20th century synthesis of Darwinian natural selection and genetical theories of inheritance (Bulmer [2003], Fisher [1953]). Finally, in the realm of philosophy of scientific explanation, Galton’s references to the deductive consequence of a statistical/mathematical law breaks the mold of typical scientific explanations that many philosophers have been focused on since the mid-20th century which concern both causal facts and relations.

According to Ian Hacking, Galton’s statistical explanation was a first of its kind in biology. Today, over a century later, statistical pattern explanations are commonplace, and not just in biology, but anywhere there are large-scale dynamical systems that are the aggregate effect of micro-scale processes, including (this is a partial list): evolving populations, thermodynamical systems, ecological systems, weather

¹ According to historian Stephen Stigler, Galton’s exposition was a better-presented version of the more important critique that Fleeming Jenkin famously challenged Darwin with Stigler ([2012]).

patterns, economies, chemical reactions, even earthquakes. Statistical pattern explanations predominately rely on limiting distributions, the normal law that Galton invoked being the most common. Limiting distributions capture the convergence of a sequence of distributions when the sequence is scaled towards an infinite value (Epps [2013]). More formally, the limiting probability distribution of some random sequence is the limiting distribution of some function of that sequence. However, not all limiting distributions generate the bell-shaped distribution pattern. There are others, including Poisson, exponential, and the power law, each revealed by a variety of statistical parameters, and each finding multiple instantiations in nature. The Poisson distribution frequently appears when one counts the number of times an event occurs per unit area or unit time and is hence applicable to a range of common events: network failures, number of website visitors, number of bubbles in glassware produced. The exponential curve appears over the waiting time for the first occurrence of an (memory-less) event: pandemics, spoilage time, compound interest. The power law appears as a functional relationship between quantities, where a relative change in one quantity results in a proportional relative change in the other quantity: magnitude of earthquakes, sizes of meteorites, losses caused by business interruptions from accidents (Newman [2017]). Despite these differences in the distribution and corresponding equations, they are all valuable to inquiry for the same reasons that Galton attributed to the normal law. They are never exactly correct in any given instance, but they are approximately true, and that is what makes them “serviceable for explanation”. To reflect this feature, Hacking calls Galton’s application of the normal law to explain a biological phenomenon a “statistically autonomous explanation” (SAE).

Despite its scientific import, philosophers of science have only recently begun to analyze the features of Galton’s statistical explanation for heredity as well as to examine statistical pattern-level explanations more generally. As regards the former, Elliott Sober ([1980]) claims that what makes Galton’s statistical explanation unique is the nature of the *explanandum*. The statistical variability enters into the explanation because “regardless of the underlying etiology” the normal distribution is a *real* phenomenon, it is subject to its own forces, and it obeys its own laws. The details of the individuals from which the population effect is comprised “are pretty much irrelevant”. Sober marks this as a feature of “population thinking” among Victorian scientists, in that it abstracts away from individuals (Sober [1980] p. 177). As regards the topic of statistical pattern-level explanations more generally, these kinds of explanations have recently received greater consideration. Prominent among the views which purport to subsume these kinds of explanations are Lange’s “really statistical” view (Lange [2013], [2016]); Strevens’ causal difference-making account (Strevens [2016]); Kaplan and Craver’s 3M/3M* mechanist view (Kaplan and Craver [2011]; Craver and Kaplan [2020]); and Ariew, Rowher and Rice’s statistical autonomous explanations view (Ariew et al. [2015], [2017]; Rice et al. [2019]).² The relevant issue for many of these accounts is whether explanations that invoke limit theorems are causal or non-causal (see Ariew, 2017 et. al. for a rebuke). But, our interest is in the following three questions, prompted by Galton’s original discussion, and made urgent by the ubiquity of statistical pattern explanations in science³:

² In what follows, we avoid prolonged discussion of the rival accounts mentioned here given that our aim is to revise the SAE view since it represents the only account of statistical pattern-level explanation which attempts to answer each of the following crucial questions (Q1-Q3).

³ This set of crucial questions (Q1-Q3) which we introduce as an adequacy benchmark for statistical pattern-level explanations generally resembles the set of questions which Batterman and Rice describe as requiring answers for the construction of what they term a “minimal model” (Batterman and Rice [2014]). The general resemblance suggests a complimentary relationship between the view presented here and their own. However, the view to be presented in this paper is focused specifically on statistical explanations oriented around limit distribution patterns,

- Q1. Why are these limit distribution patterns so common in nature?
- Q2. Why are these limit distribution patterns robust and stable, insensitive to perturbations?
- Q3. Why are most lower-scale details irrelevant to generating the limit distribution pattern?

We shall claim that while some of the extant accounts are more adequate than others in answering Q1-Q3, none of these accounts currently resolve these crucial questions. To remedy this lacuna, we argue for a modified and extended version of the SAE account. The account is both modified and extended by adopting a maximum entropy approach for generating models of the kinds of limit/probability distributions which are central in statistical pattern-level explanations (Jaynes [2003]; Frank [2009]). This achieves two important results. First, the range of the original SAE account is extended and is demonstrated to apply to a wide swath of different limit distributions. Whereas the original SAE view accounted mostly for patterns arising from the normal distribution, the modified account is applicable to cases which involve exponential, Poisson and power law distributions. Second, the maximum entropy approach for generating models of limit distributions contain the ingredients for demonstrating why lower-level details about most of the characters or events comprising the ensemble are irrelevant to the limit distribution pattern. As such, this enables the modified account to answer the crucial set of questions (Q1-Q3) referenced above, yielding a fuller and more satisfying account of statistical pattern-level explanations, and begins to capture Galton's summary of its features.

This paper will unfold as follows. In section 2, the notion of statistical autonomy as exemplified in the Galton case and Ariew et al.'s ([2015], [2017]) SAE view is unpacked. In section 3, the notion of maximum entropy as a way of modifying the account is introduced. The SAE account is then both modified and extended by adopting the maximum entropy approach for generating models of limit distributions and is demonstrated to answer the set of crucial questions (Q1-Q3) requisite for being a fuller account of statistical pattern-level explanation. In section 4, two case studies of statistical pattern-level explanations are shown to be assimilable to the modified SAE account. Section 5 concludes.

2. Statistical Autonomy and the SAE Account of Statistical Pattern-Level Explanations

2.1 The notion of statistical autonomy

Ian Hacking begins his investigation of statistical autonomy as a unique explanatory property by revisiting the Galton case. Specifically, Hacking ([1990]) provides an analysis of the nature of Galton's *explanans*, centered around making sense of Galton's remark that the statistical law in this instance is "serviceable for explanation". The key feature is its "autonomy": "one can explain something by using a statistical law only if it is in some way autonomous, and not reducible to some set of underlying causes" (Hacking [1990], p. 181). Hacking contrasts his analysis from the debate concerning explanations invoking statistical laws that were popular at the time of his writings.

First, to be a statistically autonomous explanation requires no ontological commitment to the "universe containing deeper and nonprobabilist laws that entail the statistical behaviour" (*Ibid*). To encapsulate this idea, Hacking contrasts "autonomy" with "irreducibility". Galton intended to explain what he believed was a well-confirmed phenomenon—the regular and law-like regularity of heredity patterns of families across generations. This leads to the second unique feature of Galton's statistically autonomous explanation. Many philosophers interested in statistical and probabilistic explanation are

and thus possesses a narrower yet novel scope since Batterman and Rice's view operates at a higher level of generality.

concerned with explaining events that occur with some frequency (from seldom to frequent). Galton's aim was not to explain an event but a robust phenomenon associated with heredity.

A third unique feature of Galton's statistically autonomous explanation concerns his novel interpretation of the bell-shaped distribution of characters. According to Galton's contemporaries, most notably, the astronomer-turned-social scientist, Adolphe Quetelet, the normal distribution is the limit of an aggregate effect of a lot of independent causes. Consequently, the presence of a normal distribution provides information about how each individual came to develop their traits. The mean height of a population or race represents the "constant" forces that determine a person's height, while the dispersion around the mean is a measure of the possible effects of "accidental" causes that prevents an individual from attaining its true type-characteristic. According to Galton, the problem with applying Quetelet's interpretation of the normal distribution is that it left the reason why heredity conforms to the law "totally unexplained". Now, Galton is not denying Quetelet's assertion that the normal distribution reduces to underlying causal behavior, as we acknowledged before, autonomy is not the same as irreducibility. Rather, the fact of reduction is irrelevant to the question of autonomy—why the processes of heredity maintain the whole distribution pattern—described by the normal distribution—over the course of generations. As Hacking puts it (crediting Victor Hiltz): "where Quetelet was thinking of a central tendency, and hence of the mean, Galton, always preoccupied by the exception, was thinking of the tails of the distribution, and of the dispersion. Mathematically, speaking, the mean and the dispersion are necessary and sufficient for describing the curve—co-equals as defining properties..." but Quetelet and Galton attended to them very differently. In sum, in regards to Q3, when Galton spoke about statistical laws being "serviceable for explanation", he was aware that he was explaining without references to the "host of petty independent causes" that concerned other investigators like Quetelet.

2.2 Statistical autonomous explanations

Where Hacking and Sober focus on the ontology of statistical laws in Galton's explanation, the analysis of Ariew et al. ([2015], [2017]) focuses on Galton's use of a mathematical idealization to explain a biological phenomenon: "Galton's case shows how a statistical model can be used to provide an autonomous explanation independent of any claims about what causal relationships are present in the world" (Ariew et al. [2015]). This provides the basis for their SAE account of statistical pattern-level explanations. SAEs involve the following two steps:

Step 1. Minimal Material Conditions or MMCs

On the MMC step, the real-world target system (usually, a population or ensemble of events) is made amenable to statistical analysis. Specifically, the target system is shown to satisfy the minimal material conditions (MMCs) requisite for representation as a certain stochastic ensemble. In the first step, Galton assumed that the characters in the population approximate the idealized normal distribution. The purpose of the assumption is to treat the biological phenomenon as a mathematical problem. The statistical assumption is warranted if a real-world population satisfies the following MMCs: (1) The population must feature characters that can be sampled randomly; (2) The trials by which they are sampled must be independent; (3) The trials must be sufficiently numerous (Ariew et al. [2017]). So long as the population can be shown to comport with these conditions, "it matters not whether the ensemble is composed of coin tosses, shots on target, heights of soldiers or biological characters..." the real-world population can be represented as approximating a normally distributed statistical ensemble (Ariew et al. [2017], p. 64). The minimality of these conditions is intended to be suggestive of the potentially wide scope of application for this kind of statistical representation across a range of cases. It should be noted Ariew et al.'s MMCs are

consistent with many textbook descriptions of the normal distribution or the “central limit theorem” ([Reid] 2013).

Step 2. The Application of a Statistical Law/Fact/Theorem

At this step, a statistical law/fact/theorem is applied to the statistical ensemble and the explanandum is deduced as a consequence of the law/fact/theorem’s application (Ariew et al. [2017], p. 67). In this second step, the explanandum—the stability of the normal distribution over the course of generations—is deduced as a consequence of the normal law without any appeal to the specific underlying causes or hereditary mechanisms that actually make up an instance of the phenomenon in the real biological world. Consequently, we have an explanation that features what Galton noted (in the “serviceable for explanation” quote, above). Since the explanation is an idealization, it will never be exactly correct in any one case. But, since real-world populations approximate the idealized distribution, the explanation will “most nearly express what takes place in nature generally” (Galton [1877]).

Ariew et. al. complete their analysis by addressing a further question: what is the advantage of statistically autonomous explanations over that of an explanation that cites biological causes or mechanisms?

A response is issued as follows. The advantage of Galton’s SAE explanation is that it provides “extra explanatory information that an explanation that adverts to biological causes does not apply.” To illustrate this point, Ariew et. al. ([2015]) cite Woodward discussion of “complex” or “higher order” system explanations where the explanandum involves the large-scale behavior of an ensemble that is composed of individuals that vary in their individual trajectories (Woodward is specifically talking about gas laws):

“There are a very large number of different possible trajectories of the individual molecules in addition to the trajectories actually taken that would produce the macroscopic outcome...that we would want to explain. In other words, no matter what the arrangement of the causes are, a particular ensemble level trajectory is highly likely” (Woodward [2003]).

The statistically autonomous explanation provides this sort of information because it is not limited (as causal/mechanistic explanations are) to citing the actual causal trajectories of the individuals at the lower levels. For this reason, statistically autonomous explanations are both highly general and multiply realizable (even though they never describe any actual real-world instance).

Evaluation of this view can begin by recalling the earlier set of crucial questions (Q1-Q3) and assessing how it answers them. Concerning both Q1 and Q3, an answer is in part provided by the notion of MMCs. The kinds of pattern at issue (in this case, normal distributions) are common throughout nature since the conditions requisite for generating them are minimal: all that is required for a population to approximate this distribution is that it be amenable to random, independent sampling of sufficient numerosity. Additionally, most other lower-level details pertaining to the characters which comprise the ensemble are extraneous to generating the pattern and therefore irrelevant. Given that the necessary conditions for the generation and maintenance of the pattern are so minimal and abstract, the pattern is accordingly very robust. The initial hurdle of answering to Q1-Q3 and being a complete view of statistical pattern-level explanation is thus cleared on the SAE account.

While the SAE view of Ariew et al. shows promise in answering to Q1-Q3, one ancillary issue and a serious problem arise in considering the view. First, are there SAEs which invoke limit distributions aside from normal distributions or is the Galton case and are regression cases more generally a set of special cases? For example, power law, Poisson, exponential and gamma distributions are pervasively

applied in science. Power law distributions, to take just one example, have been identified as conforming to data which captures the occurrence of extinction events, earthquakes, word frequency in books and seemingly everything else from soup to nuts (Newman [2006]). Does the SAE account have anything to tell us about these limit distributions and more broadly, what is the scope and range of this view?

Finally, the SAE view encounters a substantive problem: their description of the MMCs for generating the normal distribution and thus their account of the pattern's ubiquity in nature is likely false. Ariew et al. ([2015], [2017]) adopt the common line that the wide applicability and ubiquity of the normal distribution pattern in nature owes to something like the minimal requirements of the central limit theorem (CLT). Roughly, the CLT implies that if some random variable, X , is the sum of a large number of small and independent variables ($X_1 \dots X_n$), then, irrespective of their distributions, X will be a normally distributed variable. The CLT requires that ($X_1 \dots X_n$) must be identically and independently distributed about some mean and variance. That is, ($X_1 \dots X_n$) must all share the same probability distribution (identity condition) and ($X_1 \dots X_n$) must be mutually independent events (independence condition). This is commonly known as the "*i.i.d.*" requirement. As the number of *i.i.d.* variables for X increases, the average converges on a normal distribution. This is ostensibly what drives Ariew et al.'s claim that random independent sampling of sufficient numerosity is what allows a real-world population to approximate a normal distribution (Ariew et al. [2017]).

The CLT narrative about the commonness of the normal distribution pattern is problematic for two reasons. First, the identity condition is a restrictive one and so many cited instances of the normal distribution will fail to satisfy it. Lyon illustrates this point with the sort of case which is considered a paradigmatic instance of the normal distribution via the CLT narrative, the kind encountered in primers on statistics (Lyon [2014]). If a baker produces 100 loaves of bread and adheres to a recipe which yields loaves with an average weight of 100 grams, there will obviously be some loaves which weigh marginally more or less than others. Can the CLT explain these marginal differences? While the variables which comprise the weight of the loaf (salt, sugar, flour, yeast, water) may satisfy the independence condition, it is doubtful they satisfy identity. Plausibly, the variances of each of these variables are different; for instance, the variance of salt is not the same as the variance of water. So, this case fails to satisfy identity and by extension the *i.i.d.* requirement for the CLT. And likewise for many other such cases (Lyon [2014]).

Additionally, more careful investigation has shown that many of the cited instances of the normal distribution pattern turn out to be cases of the log-normal distribution (Lyon [2014]; Weatherall [2013]). In log-normal cases the multiplicative (not summative) product of many independent random variables is ascertained. This involves the logarithm of a random variable's probability distribution, not the random variable's probability distribution itself (Limpert et al. [2001]). It turns out that many cases of normal distributions are actually instances of these log-normal distributions (Weatherall [2013])⁴. If one of the stated goals of Ariew et al.'s SAE view is to demonstrate the wide scope of the normal distribution, then the CLT narrative they adopt likely fails in this task for the above-going reasons: first, the identity condition of the *i.i.d.* requirement is restrictive and thus not met in many cases and second, if many purported instances of the normal distribution pattern turn out to be cases of log-normal distributions, then this troubles claims about the prevalence of the normal distribution pattern in nature. The CLT type

⁴ One well-known instance of this confusion occurred in early statistical models in finance. Early on, stock prices were thought to fluctuate randomly in a way which conformed to the normal distribution. It was later discovered that this was false; the random fluctuation of stock prices better approximates a log-normal distribution (MacKenzie, [2003]; Weatherall [2013]). And similarly for many other phenomena that were originally thought to be normally distributed.

account which Ariew et al. adopt fails in its intended aim to demonstrate the wide applicability of their view since its actual domain of application is fairly narrow.

3. Modifying the SAE Account: Limit Distributions and Maximum Entropy

If limit/probability distribution patterns represent a core set of statistical patterns, then plausibly, building a fuller account of statistical pattern-level explanation warrants closer engagement with them. Indeed, these patterns act as the common denominator for statistical pattern level explanations. Once it is demonstrated that the pattern is instantiated in some population or group, certain explananda can be deduced from this and further facts. That is, to answer Q1, “why are they so pervasive in nature,” we attend to the question of how these limit distribution patterns are generated.

Our response begins with adopting a maximum entropy (ME) approach (Jaynes [2003]; Frank [2009]). On this line, a limit or probability distribution pattern is isolated via the construction of a neutral generative model where certain information constraints are set and entropy maximized for all of the information which has not been selected for.⁵ Unpacking this statement can begin with a more detailed description of the ME approach. While entropy’s conceptual origins are traceable to physics, particularly the statistical mechanics of Boltzmann and Gibbs, the concept found application in information theory.⁶ Therein, entropy is understood as randomness or uncertainty (Cover and Thomas [2006]). Information is measured, inversely, as the amount by which a message reduces entropy (Stone [2018]). More formally, let X be a discrete random variable with possible outcomes $x_1 \dots x_n$ where these outcomes occur with probability $P(x_1) \dots P(x_n)$. Entropy or H is defined as:

$$H(X) = - \sum_{i=1}^n P(x_i) \log P(x_i)$$

Where sigma expresses the sum over the possible values for the variable.

In terms of limit distributions, models with less information imply higher entropy and so to maximize entropy for some variable, one effectively minimizes the amount of prior information which is built into the distribution. The maximum entropy approach for generating limit distributions was pioneered by E.L. Jaynes who expresses his maximum entropy principle as follows:

Maximum Entropy Principle. If a certain probability distribution maximizes entropy subject to certain constraints representing our incomplete information, this is the fundamental property which justifies our use of the distribution for inference; it agrees with everything that is known but carefully avoids what is unknown (Jaynes [1990]).

Per the ME principle, the lack of prior information baked into the model is what ensures that the model is both unbiased and robust. The model only preserves information which has been encoded for by the relevant information constraints. Accordingly, the distribution which maximizes entropy is the one which contains the least amount of implicit and unwarranted assumptions.

⁵ The “neutrality” of the model owes to the fact that the model maximizes entropy, a detail which will be explained in short order.

⁶ The closeness of the information theorist’s concept of entropy and the physicist’s is a topic worthy of further exploration. However, this cryptic quote from Claude Shannon, the forefather of information theory, demonstrates the potential for cross-application: “a basic idea in information theory is that information can be treated very much like a physical quantity, such as mass or energy.” (Shannon [1985]).

To make this more intuitive, consider the following example. Suppose a hundred people are to be interviewed and asked to randomly think of a number between one and ten. You are asked to imagine the possible distributions that might result from recording their answers on a graph where the X-axis represents the ten possible responses (1-10 inclusive) and the Y-axis represents the fraction of people who submitted each number as a response. Additionally, you do not adopt any additional constraints on information. The least informative distribution for this exercise given your lack of information constraints would be one in which each number is selected with equal probability—in this case, each possible answer receives $1/10^{\text{th}}$ of the distribution. This would yield a uniform distribution or a straight line in graphical terms—each possible number value receives $1/10^{\text{th}}$ of the responses. Consequently, the uniform distribution would represent the distribution which maximizes entropy in this particular case since there is no less informative distribution for this example. This is the distribution which contains no implicit and biased assumptions.

While this general description of the ME approach is informative, a more concrete case featuring the construction of a neutral generative model is desirable. Suppose one wishes to construct a model for some process and the aim is to capture a normal distribution. The modeler sets information constraints for the mean and variance for some distribution. All other small-scale fluctuations and processes not coded for by these constraints are allowed to cancel or average out in long run. In the limit, the normal or Gaussian distribution is converged on or becomes a basin of attraction for the pattern. The upshot is that:

In terms of information, the final pattern reflects only the information content of the system expressed by the constraints on randomness; all else dissipates to maximum entropy as the pattern converges to its limiting distribution defined by its informational constraints (Van Campenhout and Cover [1981]).

A crucial insight offered by the ME approach which materializes from the foregoing description is that it capably addresses the fundamental question (i.e. Q3) about why so many of the lower-level details included in samples which yield limit distribution patterns are irrelevant.⁷ Limit distribution patterns emerge from systems which are complex and involve a plethora of small-scale subprocesses and fluctuations; what Galton referred to earlier as “petty influences.” However, the very procedure for capturing limit distributions is demonstrated to contain the ingredients for eliminating most lower-level detail which is irrelevant to the higher-level pattern. On the ME approach, the selected information constraints are preserved. But all other detail is cancelled out as entropy is maximized for all which is not coded for by the information constraints. In fact, the cancelling out or entropy maximization for this unselected information is what makes the limit distribution a basin of attraction for the process.

Additionally, the modified view avoids the implausibility which surrounds the CLT narrative which the original SAE was wrongly founded upon. The ME approach does not require the satisfaction of an *i.i.d.* requirement. By fixing both the mean and variance of the distribution while maximizing entropy for all else, we arrive at a normal distribution. And a distribution which maintains its mean and variance through repeated operations on it is a normal distribution. This fortunately avoids the problem of having to satisfy something as restrictive as the identity condition in order to generate a normal distribution.

A natural nexus thus emerges between the ME approach and the SAE account. Recall that on the SAE account, a necessary step in SAE construction was to demonstrate that the real-world ensemble satisfied the minimal material conditions for approximating a particular kind of limit distribution. In their

⁷ As S.A. Frank phrases this question, “what is the relationship between the processes which aggregate and the pattern itself?” (Frank [2009], p. 1178).

Galton case, which involved the normal distribution, these conditions were random, independent samples of sufficient numerosity (Ariew et al. [2017]). We shall argue that the coded for information constraints can play the role of the minimal material conditions for generating SAEs; indeed, the line of the ME proponents is strikingly similar to the SAE proponents. The pervasiveness of these limit or probability distribution patterns in nature owes to the minimality in what is requisite for generating them. A population's satisfying these conditions —being conformable to a modeling process whereby the relevant information constraints can be imposed— is sufficient for demonstrating that the population approximates a particular kind of limit distribution.

Returning once again to a former question we may again ask what is the scope and range of the modified SAE view? Can the view account for limit distributions beyond the normal distribution? Incorporating the ME approach into the SAE view encourages an affirmative answer. As S.A. Frank argues, most of the common limit distribution patterns can be subsumed by the ME approach (Frank [2009]). The following table (Table 1) summarizes a set of common limit/probability distribution patterns and their corresponding minimal material conditions or information constraints.⁸

Limit/Probability Distribution Pattern	Minimal material conditions/ Information Constraints
Normal	<ul style="list-style-type: none"> • Random independent sampling • Mean and variance
Poisson	<ul style="list-style-type: none"> • Random independent sampling • Discrete time • Number of events per unit area/time where N (the number of Bernoulli trials) is sufficiently large (e.g. $N > 20$) and the probability of some event is sufficiently small (e.g. $P < 0.05$).
Power Law	<ul style="list-style-type: none"> • Random independent sampling • Can we obtain a functional relationship between two quantities where a change in one quantity occasions a proportional change in another? • Scale invariant
Exponential	<ul style="list-style-type: none"> • Random independent sampling • Memoryless process • Continuous time • Sufficient number of events per unit area/time and the probability of the relevant event is sufficiently small

Table 1. Limit distributions and their corresponding minimal material conditions

The modified SAE account achieves two important results: first, by incorporating an ME approach into the view, the account can be extended to treat probability/limit distributions which go beyond the normal distribution. This is a significant result. If an important class of the statistical patterns which are deployed by scientists to account for natural patterns are the probability/limit distributions cited above, then the

⁸ Other distributions which are amenable to a maximum entropy analysis and not included in table 1 include the Gamma, LaPlacean, Pareto and Log-Normal distributions.

modified SAE view enjoys a considerable range of application. Further, the modified view describes the process by which these limit distributions are demonstrated to be applicable to real-world ensembles: this is of course the verification that these ensembles satisfy the minimal material conditions for generating these limit distribution patterns. Finally, the modified approach avoids the plausibility issues which afflicted the CLT narrative of the original view.

Aside from extending the SAE view to a range of common limit/probability distributions, incorporation of the ME approach permits fuller answers to be issued in response to the crucial set of questions (Q1-Q3):

Q1: Why are the limit distribution patterns so common in nature?

Because the requisite conditions for generating these patterns are so minimal. The minimality is evinced in the case of the normal, exponential, Poisson and power law distributions. The ME approach fills out these minimal material conditions in terms of information constraints. For example, any process which preserves information about the mean and variance of the sample but allows all else to maximally tend towards entropy admits of a normal distribution. Nature is replete with processes that are conformable to this kind of analysis, ranging from shots at a target to the heights of Scottish soldiers as Ariew et al. ([2015], [2017]) observe.

Q2: Why are these limit distribution patterns robust?

Because the higher-level limit distribution pattern is insensitive to most lower-level details. This accounts for the stability of these patterns under perturbations (Frank [2009]). Since these patterns are minimally generated, coding for only a few abstract, statistical properties, they remain stable under interventions.

Q3: Why are most lower-level details irrelevant to generating the limit distribution pattern?

Per the ME approach, these limit/probability distribution patterns are generated from neutral generative models where only a few properties are selected for as information constraints. Generative models are models of the conditional probability of some observed variable given a target variable. The neutrality issues from the fact that these models make no further assumptions about the distribution they seek to capture; i.e. they maximize entropy for all variables external to the observed and target variables (Frank [2016]). Entropy is maximized for all other small-scale processes and details, smoothing out these details in the aggregate. In the language of the SAE account, satisfaction of the minimal material conditions is sufficient to generate the pattern and all else is extraneous. The very ME process of constructing a neutral generative model for the limit distribution pattern contains the ingredients for discarding most lower-level details.

To sum up, the modified SAE account which is merged with the ME approach for generating limit/probability distributions achieves two primary results: first, the SAE view is shown to include a wider range of application, subsuming many of the core limit/probability distribution patterns which fund statistical pattern-level explanations while avoiding earlier problems. Second, the modified SAE account provides direct and substantive answers to the crucial set of questions (Q1-Q3) which are requisite for an adequate account of statistical pattern-level explanation.

4. The Modified SAE Account in Application: Case Studies

The motivation for investigating limit theorems as an explanatory tool is that the statistical patterns are, as the questions we posed about them, common in nature, robust and stable –even when the lower-level details of the ensemble are varied. In the following section we investigate two case studies to demonstrate

how the modified SAE accounts for these features of limit theorem explanations. We chose one instance that is well-discussed in the philosophy of biology literature, the application of the central limit theorem to investigate evolution by genetic drift, and one that is not so common: the investigation of the frequency of earthquakes using the power law.

4.1 Case study 1: Genetic drift

Genetic drift's explanatory contributions in population biology are multifarious. These contributions include why trait frequencies in a population depart from expectation values (where these values are determined by selection pressures); why these departures are greater in smaller as opposed to larger populations; and why trait frequencies within a population featuring no mutations, migration or selection pressures still change dynamically.

Debates over the nature of drift and how to differentiate this concept from selection in population biology remain live. On one approach, drift is regarded as sampling error (Beatty [1984]; Plutynski [2005]). On another set of views, drift is understood as a causal process or an aggregation of causal forces (Sober [1984]; Stephens [2004]; Millstein [2006]; Gildenhuis [2009]). Alternatively, Walsh et al. ([2002]) and Matthen ([2010]) argue for a mathematical/statistical characterization of a certain kind of drift which is meant to explain departures from expectation values of trait frequency. It is this view which we shall argue admits of a modified SAE interpretation.

Trait fitness is essential to explaining changes in biological populations. The distributions of trait fitness are taken to both predict and explain structural changes in a population undergoing natural selection (Walsh [2015], p. 477). A trait's fitness is equivalent to its tendency to increase or decrease its relative frequency in a population. The concept of genetic drift enters the picture in accounting for departures from expectation value, where these values are determined by selection pressures. Of crucial importance is the Hardy-Weinberg law which is summarized as follows:

The Hardy-Weinberg Law says that in infinite populations (of diploid organisms) there is no change in gene frequencies when there is no variation in gene fitnesses. But natural populations are finite in size; often they are small. In finite populations there will always be some non-negligible chance that trait frequencies will diverge from expectation. (Walsh et al. [2002], p. 456).

The divergence from expectation value can be accounted for by genetic drift in something like the following way:

...in these cases what happens is that the distribution of fitnesses in the population yields a prediction concerning the way in which a population will change. Drift is manifested as a difference from the outcome predicted by the fitnesses in the population. The law of large numbers tells us that the likelihood of significant divergence from these predictions is an inverse function of the size of the population. The small size of a population increases the chances of error. (Walsh et al. [2002], p. 459).

In order to examine how drift is isolated from selection effects, consider the following simple equation for trait fitness or w_i . The fitness w of some trait i can be modeled as follows (Gillespie [2004]):

$$w_i = \mu_i - \sigma_i^2/N$$

Where μ_i is the mean reproductive output of i individuals, σ_i^2 is the variance in reproductive output of i individuals and N is the population size parameter. Variance is often associated with "genetic drift" or the

total number of organisms in some population which either fail to reproduce or perish. In stochastic terms, per the Hardy-Weinberg law, the amount of variation turns out to be inversely proportional to the size of our population —the smaller the population, the greater the variation is expected to be and vice versa (Orr [2007]). Thus, drift can be eliminated from this equation by scaling the population parameter (N) towards an infinite value. As N approaches an infinite value, variance converges on zero. This allows selection effects or mean reproductive output to be isolated from drift.

How might this conception of genetic drift be enlisted in an explanation? Consider two populations (P1 and P2) which share the following initial conditions: (1) They are both of finite but different sizes; (2) They share the same traits and these traits have identical initial frequencies; (3) They both are expected to undergo the same selection pressures and there are no mutations or migrations. We may ask why is it that we expect to see greater departures from expectation value of trait frequencies in the smaller population, P1, than the larger population, P2? What ultimately explains the difference in P1 and P2 with respect to their expectation values is the kind of drift mentioned above: we expect greater drift in P2 than P1 owing to the respective differences in population size.

For clarity's sake, the kind of drift case mentioned above can be analogized to a coin case as follows. Let P-alpha be analogous to a run of coin tosses in which a fair coin is flipped ten times, with results tallied. Let P-beta be analogous to a trial run of coin tosses in which we flip a fair coin one hundred times. Why do we expect to see a larger divergence from expectation value (where this divergence is taken to be analogous to drift) in P-alpha relative to P-beta? An explanation can be provided which invokes something like the law of large numbers, which we define informally as expressing that the average of the results of trial runs of an experiment will likely come closer to the expectation value as trial runs are increased (Walsh et al. [2002]). Effectively, the law of large numbers tells us that the likelihood of greater divergence from expectation value is an inverse function of the size of population (Crow and Kimura [1970]). In the coin case, this can be demonstrated by conducting one hundred trials of P-alpha type tosses and obtaining the average. This can then be compared with the results of one hundred trials of P-beta type tosses, yielding a corresponding average. Given the comparative size of P-alpha to P-beta as well as the law of large numbers, we expect to see greater divergence (drift) in P-alpha relative to P-beta.

Returning to the biological populations of the initial drift case, (P1 and P2), this kind of explanation can be demonstrated to admit of a modified SAE interpretation as follows. First, it is assumed that the trait distributions in both populations approximate a certain kind of limit distribution: namely, a normal distribution. Establishing this fact requires checking that both populations satisfy the MMCs. Notice that both populations can be thought of as processes involving random, independent sampling. These ensemble or population features are what cement the analogy between the expected drift case and the toy coin case just described. Indeed, it is this very underlying assumption that drove the plausibility of the sampling error interpretation of drift. In terms of the maximum entropy approach, the Gillespie equation, which expresses trait fitness, codes for both mean and variance. Here, the statistically autonomous character of the explanation manifests: most lower-level micro-causal details of the populations are ignored in thinking of them as approximations of normally distributed ensembles.⁹ Once this step is completed, we then apply a statistical law/fact/theorem to our target systems represented as ensembles. In this case, the amount of drift or variance can be determined via inspection of the variance parameter in the foregoing equation. In P1, the smaller population which ipso facto includes more variance, the amount of expected drift is less than P2 which accounts for the greater departure from expectation value in P1 versus P2. Per the law of large numbers as it relates to the Hardy-Weinberg law,

⁹ The notion of drift-based explanations as ignoring most details is shared in Matthen's characterization of drift explanations as statistically abstractive explanations (Matthen [2009]).

the smaller size of P1 helps to explain the greater amount of drift in that population relative to P1. Given P1’s smaller size, it is further away from Hardy-Weinberg equilibrium than P2. A modified SAE interpretation thus materializes in this genetic drift case.

4.2 Case study 2: Earthquakes

Suppose at some geophysical region, R, many earthquakes of smaller magnitude are observed relative to earthquakes of larger magnitudes. A natural question is to inquire why this is so. One means of response involves inspection of the geophysical specifics which pertain to R in an effort to reveal some underlying mechanism. Another more statistical and pattern-oriented approach involves reference to the Gutenberg-Richter law (GRL) in seismology. Per GRL, earthquake magnitudes and the number of earthquakes at a given region approximate a power-law relationship (Bak [1996]). This can be formally expressed as follows:

$$\log_{10} N = a - bM$$

Where *N* is the number of events with a magnitude greater than or equal to *M*. Both *a* and *b* are constants. “*a*” is a parameter representing total seismicity and “*b*” or the “*b* value” is a scaling parameter which relates the number of large events to small ones. In seismically active regions, the *b* value is often near 1 (Bak [1996]). This implies that for earthquakes of magnitude 3, there will be ten times as many quakes with a magnitude of 2 and 100 times as many quakes with a magnitude of 1. Finally, a logarithm is deployed since a histogram of quantities which are power law distributed can be graphically converted into a straight line at logarithmic scales; e.g. a log of the frequency declines linearly with a log of the size (Newman [2006]). Somewhat astonishingly, GRL has been confirmed as an empirical, seismological law for up to 8 orders of magnitude on the Richter scale (Smith [1981]).

Understanding the significance of GRL necessitates further examination of power laws. A power law expresses a functional relationship between two quantities where a relative change in one quantity occasions a proportional, relative change in the other quantity. This change is independent of the size of the quantities. A power law can be expressed in generic form as follows (Frank [2009]):

$$\text{Power Law in generic form: } Y = k X^a$$

Where *X* and *Y* are the two quantities, *a* is the law’s exponent and *k* is some constant. Additionally, power laws feature the property of scale invariance. This means that changes in frequencies remain constant at all scales. Another way of putting the point is that power laws have fractal properties (Mitchell [2009]). This special feature accounts for both the robustness and stability of power laws. Power laws graphically feature a characteristic long tail (see Figure 1 below).

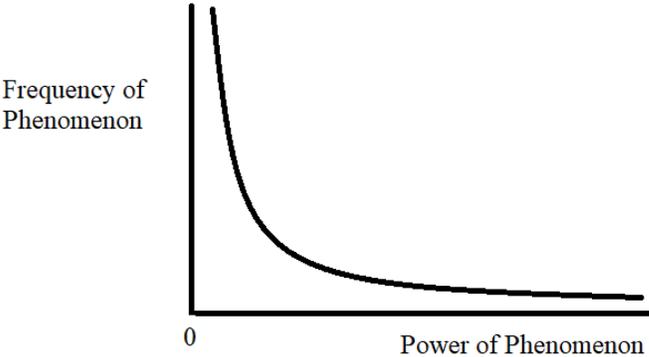


Figure 1. Generic graphical representation of a power law curve

Returning to our earlier explanandum –viz. for some region R, why are there many more earthquakes of smaller magnitudes than earthquakes of larger magnitudes? –we can provide a partial statistical pattern-level explanation as follows. Per the GRL, for any geophysical region, earthquake magnitudes and frequencies can be shown to follow a power law distribution. And the power law distribution expressed by GRL implies that relative to earthquakes of larger magnitude, there will be many more earthquakes of smaller magnitudes. Notice that GRL posits that frequencies stand in an inversely proportional relationship to magnitude. Since the GRL is an empirical law, this applies to our region of interest, R. Thus, there are many more seismic events of smaller magnitudes than larger ones at R.

Of interest for our purposes is whether this partial statistical pattern-level explanation can benefit from treatment as a modified SAE. Does the maximum entropy approach have any contribution to make in enhancing this explanation? In the earthquake case, a statistical pattern which ranges over a diverse set of phenomena –i.e. many regions of diverse geophysical detail– is captured by an empirical law or the GRL. Accounting for the generation of this pattern –namely, that earthquakes frequencies and magnitudes are always power law distributed– can be treated by an SAE which demonstrates that this set of phenomena satisfies the MMCs for generating a power law distribution. This might involve the following two MMCs: (1) That there be a functional relationship between two quantities; (2) That this relationship exhibits scale invariance. Notice that in the earthquake case, per the GRL, we have two quantities (earthquake magnitude and frequency) which stand in such a functional relationship. Also, this relationship is scale invariant.

The contribution of maximum entropy and information constraints enters in the following way. Answering the question of why power laws are so common in nature, S.A. Frank explains as follows:

Power laws arise by aggregation over many multiplicative processes, such as growth. Many processes in nature apply a recursive repetition of a simple multiplicative transformation, with some randomness...Aggregation over a random multiplicative process often erases all information except the average logarithm of the multiplications. (Frank [2016]).

Setting the information constraints to code for two quantities (e.g. earthquake frequency and magnitude) and allowing entropy to be maximized for all other random fluctuations and processes effectively produces the power law distribution. This is precisely the kind of constraining of information which is operative in our seismological case. Other details which are irrelevant to producing the power law pattern are allowed to cancel out and are thus irrelevant. The pattern is common since, per Frank, the kinds of multiplicative processes which involve recursive repetition are ubiquitous in nature. And the pattern's generation being dependent on so few constraints in addition to the property of scale invariance accounts for both its robustness and stability.

5. Conclusion

The central task of this paper has been to modify the original SAE account in order to provide a fuller account of statistical pattern-level explanations. The modified account, which adopts a maximum entropy approach to understanding limit distributions, has been demonstrated to answer the crucial set of questions (Q1-Q3) which were introduced as an adequacy benchmark on statistical pattern-level explanations. Further, one advantage of the modified account has been the considerably greater range of application this view possesses over the original SAE view. Adoption of the maximum entropy approach thus enables the modified view to treat a larger set of limit distributions.

One can begin to appreciate the advantage of the modified account's increased range by considering the way in which limit distributions feature prominently in model-based explanations of complex systems such as the earthquake case of 4.2. The increasing focus among scientists on large scale dynamical systems that are pervaded by complexity prioritizes a solution of the following modeling problem. Complex systems include the following three features: (1) they include an abundance of potentially relevant mechanisms across many length scales; (2) these systems often have fuzzy or indeterminate boundaries; (3) many system-level processes are laden with contingency or chaos (Harte [2011]). A further complication is that in complex systems, macro-scale behavior is often not predictable from the micro-states of the system (Bak [1996]). As such, an attempt to parameterize all of the potential mechanisms relevant to capturing the behavior of such a system would be folly.

How then does one grapple with the task of modeling complex systems? A solution often takes the form of attempting to identify a macro-scale pattern of the system which is often by necessity both abstractive and statistical (Mitchell [2009]). A subset of these patterns are explanatory, orienting around common probability/limit distributions where the phenomenon of interest can be accounted for by the fact that the target system satisfies the requisite conditions for approximating these distributions.¹⁰ This proved to be the breakthrough insight by Galton – patterns of intergenerational heredity could be accounted for by the fact that his target system (populations) satisfied the conditions for approximating normal distributions and thus feature a reversion to the mean in the distribution (i.e. heredity) of traits. Thus, a viable account of statistical pattern-level explanation demands closer engagement with these limit/probability distributions. The modified account embraces this challenge and thus is especially well-suited to subsume these kind of macro-level pattern explanations which are increasingly the coin of the realm in the scientific modeling of complex systems.

A further desideratum of a viable account is that it should provide answers to the set of crucial questions (Q1-Q3) in order to demonstrate the explanatory force of these limit distributions. In part, this is what makes the account explanatory rather than merely a form of curve fitting. The commonness of these distributions, their robustness and their insensitivity to most lower-level detail are important properties that must be accounted for if the explanatory quality of these distributions is to be made manifest.

The SAE account of Ariew et al. ([2015], [2017]) represented a promising view inasmuch as it correctly identified the relevant set of questions for a viable account of statistical pattern-level explanation. However, this view foundered in treating only one kind of limit distribution and relied upon a false CLT-driven narrative which produced only false answers to the crucial questions. The modified SAE view represents a superior alternative. By adopting the maximum entropy approach for generating limit distributions, the modified account ranges over a large swath of limit distributions, thus expanding its scope for statistical pattern-level explanations. Further, the modified view provides satisfying answers to the crucial set of questions, thus meeting a necessary condition for being an adequate view of statistical pattern-level explanation. The heightened emphasis on complex systems throughout science and the mostly statistical methods used to understand them calls out for an adequate view of statistical pattern-level explanation. The modified view stands as a useful tool for gaining insight into a rapidly growing area of scientific explanation and methodology.

Acknowledgements

¹⁰ “Subset” here is an important qualifier since certainly not all macro-scale patterns are explanatory. Admittedly, some patterns are simply phenomenological and parsing these from genuine explanatory patterns is an adequacy condition for a view of statistical pattern-level explanation or so we have argued.

The authors wish to express their gratitude to S.A. Frank as well as two anonymous reviewers for helpful comments on the draft.

Author Names and Department: Travis Holmes and Andre Ariew, Department of Philosophy, University of Missouri, Columbia, MO 65201.

Email: tlhgyf@mail.missouri.edu

References

- Ariew, A., Rice, C. and Rowher Y. [2015]: ‘Autonomous-Statistical Explanations and Natural Selection,’ *The British Journal for the Philosophy of Science*, **66**, pp. 635-658.
- Ariew, A., Rowher, Y., and Rice, C. [2017]: ‘Galton, Reversion and the Quincunx: The Rise of Statistical Explanation,’ *Studies in History and Philosophy of Biological and Biomedical Sciences*, **66**, pp. 63-72.
- Bak, P. [1996]: *How Nature Works: The Science of Self-Organized Criticality*. New York: Copernicus Press.
- Batterman, R.W. and Rice, C. [2014]: ‘Minimal Model Explanations,’ *Philosophy of Science*, **81**, pp. 349-76.
- Beatty, J. [1984]: ‘Chance and Natural Selection,’ *Philosophy of Science*, **51**, pp. 183-211.
- Bulmer, M. [2003]: *Francis Galton: Pioneer of Heredity and Biometry*. Baltimore: Johns Hopkins Press.
- Cover, T. and Thomas, J. [2006]: *Elements of Information Theory*, New York: Wiley.
- Craver, C.F. and Kaplan, D. [2020]: ‘Are More Details Better? On the Norms of Completeness for Mechanistic Explanations,’ *British Journal for the Philosophy of Science*, **71**, pp. 287-319.
- Crow, J.F. and Kimura, M. [1970]: *An Introduction in Population Genetics Theory*. New York: Harper and Row.
- Fisher, R. [1953]: ‘The Expansion of Statistics,’ *Journal of the Royal Statistical Society, Series A (General)*, **116**, pp. 1-10.
- Frank, S.A. [2009]: ‘The Common Patterns of Nature,’ *Journal of Evolutionary Biology*, **22**, pp. 1563-1585.
- Frank, S.A. [2016]: ‘The Invariances of Power Law Size Distributions,’ *F1000 Research*, **5**, 2074.
- Galton, F. [1877]: ‘Typical Laws of Heredity,’ *Proceedings of the Royal Institute of Great Britain*, **8**, pp. 282-301.
- Gildenhuys, P. [2009]: ‘An Explication of the Causal Dimension of Drift,’ *The British Journal for the Philosophy of Science*, **60**, pp. 521-555.
- Gillespie, J.H. [2004]: *Population Genetics: A Concise Guide*. Baltimore: Johns Hopkins University Press.
- Hacking, I. [1990]: *The Taming of Chance*, Cambridge: Cambridge University Press.

- Harte, J. [2011]: *Maximum Entropy and Ecology: A Theory of Abundance, Distribution, and Energetics*. Oxford: Oxford University Press.
- Jaynes, E.T. [1990]: ‘Probability Theory as Logic,’ in *Maximum Entropy and Bayesian Methods*, (ed.) P. Fougere. Kluwer: Dordrecht.
- Jaynes, E.T. [2003]: *Probability Theory: The Logic of Science*. Cambridge: Cambridge University Press.
- Lange, M. [2013]: ‘What Makes a Scientific Explanation Distinctively Mathematical?’ *The British Journal for the Philosophy of Science*, **64**, pp. 485-511.
- Lange, M. [2016]: *Because Without Cause: Non-causal Explanations in Science and Mathematics*, Oxford: Oxford University Press.
- Limpert, E., Stahel, W.A., and Abbt, M. [2001]: Log-normal Distributions Across the Sciences: Keys and Clues, *Bioscience*, *51*, pp. 341-352.
- Lyon, A. [2014]: ‘Why are Normal Distributions Normal?’ *The British Journal for the Philosophy of Science*, *65*, pp. 621-649.
- MacKenzie, D. [2006]: *An Engine Not a Camera: How Financial Models Shape Markets*. Cambridge: MIT Press.
- Matthen, M. [2009]: ‘Drift and Statistically Abstractive Explanation,’ *Philosophy of Science*, **76**, pp. 464-487.
- Matthen, M. [2010]: ‘What Is Drift? A Response to Millstein, Skipper and Dietrich,’ *Philosophy and Theory in Biology*, **2**, e102.
- Millstein, R.L. [2006]: ‘Natural Selection as a Population-Level Causal Process,’ *The British Journal for the Philosophy of Science*, **57**, pp. 627-653.
- Mitchell, M. [2009]: *Complexity: A Guided Tour*. Oxford: Oxford University Press.
- Newman, M.E.J. [2006]: ‘Power Laws, Pareto Distributions and Zipf’s Law,’ *Contemporary Physics*, **46**, pp. 323-351.
- Newman, M.E.J. [2017]: ‘Power Law Distribution,’ *Significance*, **14**, pp. 10-11.
- Orr, H.A. [2007]: ‘Absolute Fitness, Relative Fitness, and Utility,’ *Evolution*, **61**, pp. 2997-3000.
- Pierce, J.R. [1980]: *An Introduction to Information Theory: Symbols, Signals and Noise*, New York: Dover.
- Plutynski, A. [2005]: Parsimony and the Fisher-Wright Debate,’ *Biology and Philosophy*, **20**, pp. 697-713.
- Reid, H. [2013]: *Introduction to Statistics*. Los Angeles: Sage.
- Rice, C., Rowher Y. and Ariew, A. [2019]: ‘Explanatory Schema and the Process of Model Building,’ *Synthese*, *196*, pp. 4735-4757.
- Smith, W.D. [1981]: ‘The b-value As an Earthquake Precursor,’ *Nature*, **289**, pp. 136-139.
- Sober, E. [1980]: ‘Evolution, Population Thinking and Essentialism,’ *Philosophy of Science*, **47**, pp. 350-383.

- Sober, E. [1984]: *The Nature of Selection*. Cambridge: MIT Press.
- Stephens, C. [2004]: 'Selection, Drift and the Forces of Evolution,' *Philosophy of Science*, **71**, pp. 550-570.
- Stigler, S.M. [1990]: *The History of Statistics: The Measurement of Uncertainty Before 1900*. Cambridge: Harvard University Press.
- Stigler, S.M. [2012]: 'Studies in the History of Probability and Statistics, L: Karl Pearson and the Rule of Three,' *Biometrika*, **99**, pp. 1-14.
- Stone, J.V. [2018]: *Principles of Neural Information Theory: Computational Neuroscience and Metabolic Efficiency*, Sheffield, U.K.: Sebtel Press.
- Strevens, M. [2016]: 'How Idealizations Provide Understanding,' *Explaining Understanding New Perspectives from Epistemology and Philosophy of Science*, pp. 37-49.
- Van Campenhout, J.M., and Cover, T.M. [1981]: 'Maximum Entropy and Conditional Probability,' *IEEE Trans. Inf. Theory*, **27**, pp. 483-489.
- Walsh, D.M. [2015]: 'Variance, Invariance and Statistical Explanation,' *Erkenntnis*, **80**, pp. 469-489.
- Walsh, D.M., Lewens, T. and Ariew, A. [2002]: 'The Trials of Life: Natural Selection and Random Drift,' *Philosophy of Science*, **69**, pp. 452-473.
- Weatherall, J.O. [2013]: *The Physics of Wall Street*. Boston: Mariner Books.
- Woodward, J. [2003]: *Making Things Happen: A Theory of Causal Explanation*. Oxford: Oxford University Press.

